



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 113237**

**TO: Robert Kelly**  
**Location: rem-2c55**  
**Art Unit: 1632**  
**February 5, 2004**

*2CR*

**Case Serial Number: 09971773**

**From: P. Sheppard**  
**Location: Remsen Building**  
**Phone: (571) 272-2529**

**sheppard@uspto.gov**

### **Search Notes**

**THIS PAGE BLANK (USPTO)**

STIC-Biotech/ChemLib

113237

From: Kelly, Robert  
Sent: Friday, January 30, 2004 6:03 PM  
To: STIC-Biotech/ChemLib  
Subject: 09/971,773

With regard to Application No. 09/971,773, please run the following sequence searches:

Search for SEQ ID NOS: 1, 2, 23 and 24, with results of 80% or greater sequence homology

Also, search these same sequences allowing for at least 1 insertion, deletion, substitution, or addition

Thanks!

Robert M. Kelly, Ph.D.  
Room 2C55, Remsen Bldg.  
(571) 272-0729

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: February 2, 2004, 10:19:28 ; Search time 7369.83 Seconds  
(without alignments)  
11146.322 Million cell updates/sec

Title: US-09-971-773-1  
Perfect score: 2008  
Sequence: 1 aacagaaactattttctg.....gctgtgcctcaagcccatg 2008

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 2889711 seqs, 2045481386 residues  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl :

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.ph.\*
- 7: gb.pl.\*
- 8: gb.pr.\*
- 9: gb.ro.\*
- 10: gb.sts.\*
- 11: gb.un.\*
- 12: gb.vi.\*
- 13: gb.vl.\*
- 14: gb.vr.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.on.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vl.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2008	100.0	2008	6	BD168516 Cells pro
2	1809.4	90.1	2976	10	BC010666 Mus muscu
3	1804.6	89.9	2176	10	AB025198 Mus muscu
4	1584	78.9	1728	6	BD168517 Cells pro
5	1575.2	78.4	2796	9	HSAS39535 Homo sapi
6	1575.2	78.4	2898	9	HSY17976 Homo sapien
7	1575.2	78.4	2902	9	HSAS39536 Homo sapi
8	1575.2	78.4	2992	9	HSY17977 Homo sapien
9	1575.2	78.4	3186	9	HSY17978 Homo sapien
10	1575.2	78.4	3280	9	HSY17979 Homo sapien
11	1575.2	78.4	3568	9	HSAS36055 Homo sapi
12	1575.2	78.4	3666	9	HSAS36053 Homo sapi
13	1575.2	78.4	3772	9	HSAS36054 Homo sapi
14	1575.2	78.4	4196	9	HSAS36056 Homo sapi
15	1572.8	78.3	2100	6	EL5725 Human mRNA
16	1571.8	78.3	2002	9	D89289 Homo sapien
17	1568	78.1	2100	6	ARI170077 Sequence
18	1514.8	75.4	1759	6	AX662735 Sequence
19	1514.8	75.4	1759	6	BD087745 Plant cel
20	1509.6	75.2	1836	4	D86723 Porcine mRN
21	1474.2	73.4	1728	6	ARI170074 Sequence
22	1474.2	73.4	1728	6	E14720 Procline mRN
23	1469.4	73.2	1728	4	AF247186 Bos tauru
24	1256.4	62.6	2682	9	BC025385 Homo sapi
25	1090.4	54.3	2129	5	XLA514872 Xenopus l
26	983	49.0	1850	9	AF052088 Homo sapi
27	976.4	48.6	979	6	BD168521 Cells pro
28	893.2	44.5	979	6	BD168522 Cells pro
29	881	43.9	1017	6	E43925 Antibody re
30	809.8	40.3	1772	9	AB049740 Homo sapi
31	599.8	29.9	699	6	E43926 Antibody re
32	446.4	22.2	110000	2	AC129041_1 Continuation (2 of
33	446.4	22.2	110000	2	AC129041_2 Continuation (3 of
34	446.4	22.2	150709	2	AC134206 Rattus no
35	446.4	22.2	237954	2	AC128532 Rattus no
36	441.8	22.0	844	11	BV033231 S212P6677
37	441.8	22.0	171073	2	AC132342 Mus muscu
38	420.4	20.9	1262	9	HSAS14325 Homo sapi
39	420.4	20.9	1396	9	HSAS14324 Homo sapi
40	367.6	18.3	1860	3	AF441264 Drosophil
41	367.6	18.3	2764	3	AY051451 Drosophil
42	336.8	16.8	2860	3	AK113018 Ciona int
43	305.6	15.2	394	6	BD071198 Secreted
44	302	15.0	9196	6	BD168518 Cells pro
45	300.4	15.0	1147	9	AB049828 Homo sapi

ALIGNMENTS

RESULT 1  
BD168516  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

BD168516  
Cells producing antibody composition.  
BD168516  
WO 0231140-A/1.  
Cricetulus griseus (Chinese hamster)  
Cricetulus griseus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
Cricetulus.  
1 (bases 1 to 2008)  
Kanda, Y., Sato, M., Nakamura, K., Uchida, K., Shinkawa, T., Yamane, N.,  
Hosaka, E., Yamasaki, M. and Hanai, N.

2008 bp DNA linear PAT 17-JAN-2003



Qy	1741	GGAAAAACAGGCCTGTACCCCTTCCTCAAAAGTCCGAGAGAAGATAGAAAACAGTCAAAATAC	1800
Db	1741	GGAAAAACAGGCCTGTACCCCTTCCTCAAAAGTCCGAGAGAAGATAGAAAACAGTCAAAATAC	1800
Qy	1801	CCTACATATCCTGAAAGCTGAAAAATAGAGATGGAGTGTAAAGAGATTAAACAACAGAAATTTA	1860
Db	1801	CCTACATATCCTGAAAGCTGAAAAATAGAGATGGAGTGTAAAGAGATTAAACAACAGAAATTTA	1860
Qy	1861	GTTTCAGACCATCTCAGCCAAGCAGAGAACCAGCCAGACTAAACATATGTGGTTCATTGTGACGACAT	1920
Db	1861	GTTTCAGACCATCTCAGCCAAGCAGAGAACCAGCCAGACTAAACATATGTGGTTCATTGTGACGACAT	1920
Qy	1921	GCTCCGGAACAAGACAAGTGGGAAACCTCAGATGCTGCACCTGGTGGAAACGCCTCTTTGT	1980
Db	1921	GCTCCGGAACAAGACAAGTGGGAAACCTCAGATGCTGCACCTGGTGGAAACGCCTCTTTGT	1980
Qy	1981	GAAGGGCTGCTGTCGCCCTCAAGCCCATG	2008
Db	1981	GAAGGGCTGCTGTCGCCCTCAAGCCCATG	2008

RESULT 2	
BC010666	
LOCUS	
DEFINITION	2976 bp mRNA linear ROD 16-APR-2003 Mus musculus fucosyltransferase 8, mRNA (cDNA clone MGC:11418 IMAGE:3594582), complete cds.
ACCESSION	BC010666
VERSION	BC010666.1 GI:14715012
KEYWORDS	MGC.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE	1 (bases 1 to 2976)

Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [ang@bcm.tmc.edu](mailto:ang@bcm.tmc.edu)  
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,  
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>  
Series: IRAK1.P: 16 Row: 1 Column: 5  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 8393370.

FEATURES  
source  
Location/Qualifiers  
1..2976  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="MG0:11418 IMAGE:3594582"  
/tissue\_type="Mammary tumor, C3(1)-Tag model. Infiltrating  
ductal carcinoma. 5 month old virgin mouse."  
/clone\_lib="NCI CGAP\_Mam6"  
/lab\_host="DH10B"  
/notes="Vector: pCMV-SPORT6"  
1..2976  
/gene="Fut8"  
/db\_xref="locusID:53618"  
/db\_xref="MGI:1858901"  
584..2311  
/codon\_start=1  
/product="fucosyltransferase 8"  
/protein\_id="AAH10666.1"  
/db\_xref="GI:14715013"  
/db\_xref="locusID:53618"  
/translacion="NRWATGSRWIMLILFAGWTLFYIGHLVRDNDHPDHSRRLS  
KILALERLKNQEDLRNMAESLRIPEDIGQATGRVRLVEQIYKAEKQENYKK  
QARNGKXKHLELRRIENGAEKLFQSELUKHLGNEQLQVHDAELDLGLHGGH  
RSIMYDLSDQDAGWREKADLTSLVQRITTYLQNPDKCSKARKLVCKNKKG  
GCGHLHVYCFMAYGTQITLLESONWRVATGWETVFRPVSCTCTDRLGSLTGH  
WSGENDKIQVVELPIVDLHPRPYLPALVPEDLADRLRVHGDPAVMVVSQFYK  
LIRPQWLEKLEAEATKLGKHPVIGVHVRETDKVGTEAAFPPIEYVMVVEEHFOL  
LARNQVKKRVYLATDPTLLKEAKTKSYNEFISDNISISAGLHRRYTESLGV  
ILDHFLSQADFLVCFSSQVCAVEIMQTLHPDASANFHSLDDIYYPGGQNAHNQI  
AYVPHKPRTEIEIPNEPGDIIGVAGNHWDGYSKGNRKLGTGLYPSYKVRBKIEVTV  
YTPTEAK"

BASE COUNT	909 a	626 c	650 g	781 t
ORIGIN				
Query Match	90.1%	Score 1809.4;	DB 10;	Length 2976;
Best Local Similarity	94.1%	Pred. No. 0;		
Matches 1891;	Conservative	0;	Mismatches 116;	Indels 2;
Gaps	1;			
Qy	1	AAACAGAACTTATTTTC	TGTCGTGCTAACTAGAACCCAGGTACAAATGTTTCCAAATCTT	60
Db	483	AAAAGAAACCTATTTTC	TGTCGTGCTAACTAGAACCCAGGTACAAATGTTTCCAGTCTT	542
Qy	61	TGAGCTCCGAGAGACAG	AGGAGATTGAAACTCTGAAATCGGGCATGCGCTGGTT	118
Db	543	TGAGCTCCGAGAGATAGAGGACAGATTGAAACTCTGAAATCGGGCATGCGCTGGTT	602	
Qy	119	CCTGGCGTTGGATTATG	TCAATCTTTTTCCTGGGGGACCTTATGTTTATATAGGTG	178
Db	603	CCTGGCGTTGGATTATG	TCAATCTTTTTCCTGGGGGACCTTATGTTTATATAGGTG	662
Qy	179	GTCATTTGTTTCGAGATATGACCACTTACAGAGAGACTCTCCAGATTC	238	
Db	663	GTCATTTGTTTCGAGATATGACCACTTACAGAGAGACTCTCCAGATTC	722	
Qy	239	TTGCAAGCTCGAGCGCTTAAACACAAATGAAGACTTGAGGAGAAATGGCTGAGTCTC	298	
Db	723	TTGCAAGCTTGAACGC	TAAACACAAATGAAGACTTGAGGAGAAATGGCTGAGTCTC	782
Qy	299	TCCGAATACCAAGGCGCTTATGATCAGGGGACAGTACAGGAAGAGTCCGCTGTTTATAG	358	

|||||  
783 TCCGAATACAGAGGCCCATTTGACAGGGGACAGCTACAGGAGAGTCCGTGTTTAG 842  
QY  
359 AAGAAACAGCTTCTTAAGGCCAAGAACAGATTGAAATTTACAAGAAACAAGCTAGGAATG 418  
Db  
843 AAGAAACAGCTTCTTAAGGCCAAGAACAGATTGAAATTTACAAGAAACAAGCTAGGAATG 902  
QY  
419 ATCTGGGAAAGGATCATGAAATCTTTAAGGAGGAGGATTGAAATGGAGCTAAAGAGCTCT 478  
Db  
903 GTCTGGGAAAGGATCATGAAATCTTTAAGGAGGAGGATTGAAATGGAGCTAAAGAGCTCT 962  
QY  
479 GGTCTTTTCTCAAAAGTGAATGAAGAAATTAAGAAATTAAGAAAGAAAGAAAGCTCCAAA 538  
Db  
963 GGTCTTTTCTCAAAAGGCAACTGAAGAAATTAAGCAATTTAGAAGGAAATGAACCTCCAAA 1022  
QY  
539 GACATGCAGATCAAAATCTTTTGGATTTAGGACATCATGAAAGTCTATCATGACAGATC 598  
Db  
1023 GACATGCAGATCAAAATCTTTTGGATTTTAGGACATCATGAAAGTCTATCATGACAGATC 1082  
QY  
599 TATACTACCTCAGTCAAGCAGATGGAGCAGGTGAGTGGCGGGAAGAAAGCCAAAGATC 658  
Db  
1083 TATACTACCTCAGTCAAGCAGATGGAGCAGGTGAGTGGCGGGAAGAAAGCCAAAGATC 1142  
QY  
659 TGACAGAGCTGTCTCAGCGGGAATTAATATCTGCAGATCCCAAGAGCTGCAGCAAG 718  
Db  
1143 TGACAGAGCTGTCTCAGCGGGAATTAATATCTGCAGATCCCAAGAGCTGCAGCAAG 1202  
QY  
719 CCAGAAAGCTGTATGTAATATCAAAAGGCTGTGGCTATGGATGTCACCTCCATCATG 778  
Db  
1203 CCAGAAAGCTGTGTGTAATATCAAAAGGCTGTGGCTATGGATGTCACCTCCATCATG 1262  
QY  
779 TGGTCTACTGCTTCATGTTGCTTATGGCACCAGCGAACTCATCTTTGGAATCTCAGA 838  
Db  
1263 TGGTCTACTGCTTCATGTTGCTTATGGCACCAGCGAACTCATCTTTGGAATCTCAGA 1322  
QY  
839 ATTGGCGTATGCTACTGAGAGATGGGAGATGTTGTTAGACCTGTAAGTGAGACATGCA 898  
Db  
1323 ATTGGCGTATGCTACTGAGAGATGGGAGATGTTGTTAGACCTGTAAGTGAGACATGTA 1382  
QY  
899 CAGACAGTCTGGCTCTCCACTGCACACTGTGAGGTGAGTGAGGAGCAAAAGTTTC 958  
Db  
1383 CAGACAGATCTGGCTCTCCACTGCACACTGTGAGGTGAGTGAGGAGCAAAAGTTTC 1442  
QY  
959 AAGTGGTGGAGTCCCGCTTTGAGACAGCTCCATCTCGTCTCTTACTTACCTTGG 1018  
Db  
1443 AAGTGGTGGAGTCCCGCTTTGAGACAGCTCCATCTCGGCTCTTACTTACCTTGG 1502  
QY  
1019 CTGTACCAAGAACCTTGCAGATCGACTCTGAGAGTCCATGGTGTATCTGCAGTGTGGT 1078  
Db  
1503 CTGTACCAAGAACCTTGCAGACCGACTCTTAAGAGTCCATGGTCACTGCAGTGTGGT 1562  
QY  
1079 GGGTATCCAGTTGTCAAATCTTGTATCGTCCCAACCTTGGCTGGAAAGGAAATAG 1138  
Db  
1563 GGGTATCCAGTTGTCAAATCTTGTATCGTCCCAACCTTGGCTGGAAAGGAAATAG 1622  
QY  
1139 AAGAAACACCAAGAGCTTGGCTTCAAAACATCCAGTATTGGAGTCCATGTCCAGACGCA 1198  
Db  
1623 AAGAAACACCAAGAGCTTGGCTTCAAAACATCCAGTATTGGAGTCCATGTCCAGACGCA 1682  
QY  
1199 CTGACAAAGTGGGAACAGAGCAGCTTCCATCCCATTTGAGGAATACATGGTACAGTTG 1258  
Db  
1683 CAGACAAAGTGGGAACAGAGCAGCTTCCACCCCATCGAGGATGATGGTACAGTTG 1742  
QY  
1259 AAGACATTTTTCAGCTTCTCGAACCCAGAAATGAAGTGAATGAAGAGTGTATCTGG 1318  
Db  
1743 AAGACATTTTTCAGCTTCTCGAACCCAGAAATGAAGTGAATGAAGAGTGTATCTGG 1802  
QY  
1319 CCACTGATGACCTCTTCTTTTAAAGGAGGCAAGACAAAGTACTCCAAATATGAATTTA 1378  
Db  
1803 CTACTGATGATCTTACTTCTTTTAAAGGAGGCAAGACAAAGTACTCCAAATATGAATTTA 1862  
QY  
1379 TTATGATTAACCTCTATTTCTTGGTCAAGTGGATACAAACCGATACACAGAAATTCAC 1438  
|||||

Db 1863 TTATGATTAACCTCTATTTCTTGGTCAAGTGGAGTACACAAATCGGTACACAGAAATTCAC 1922  
QY 1439 TTCGGGGCGTGAATCCTCGATATACACTTTCTCTCCAGGCTGACTTCTTGTGTGACTT 1498  
Db 1923 TTCGGGGTGTGATCCTCGATATACACTTTCTCTCAAGGCTGACTTCTTAGTGTGACTT 1982  
QY 1499 TTTTCATCCAGGCTGTAGGTTGCTTATGAAATCATGCAACACTGCATCTGATGCT 1558  
Db 1983 TTTTCATCCAGGCTGTAGGTTGCTTATGAAATCATGCAACACTGCATCTGATGCT 2042  
QY 1559 CTGCAAACTTCCATCTTTAGATGACATCTACTATTTTGGAGGCAAAATGCCACAAC 1618  
Db 2043 CTGCAAACTTCCATCTTTAGATGACATCTACTATTTTGGAGGCAAAATGCCACAATC 2102  
QY 1619 AGATTGCAAGTTTATCTCTCACCACCTCGAACTTAAAGAGGAAATCCCATGGAACTGGAG 1678  
Db 2103 AGATTGCTGTTATCTCTCACAACCTCGAACTGAAGAGGAAATCCAATGGAACTGGAG 2162  
QY 1679 ATATCATTTGGTGTGGTGGAAACCATTTGGAATGTTACTCTAAAGGTGTCAACAGAAAC 1738  
Db 2163 ATATCATTTGGTGTGGTGGAAACCATTTGGAATGTTACTCTAAAGGTGTCAACAGAAAC 2222  
QY 1739 TAGGAAAACAGGCTGTACCTTCTACAAAGTCCGAGAGAGATAGAAACAGTCAAT 1798  
Db 2223 TTGAAAACAGGCTTATATCTCTTACAAAGTCCGAGAGAGATAGAAACAGTCAAT 2282  
QY 1799 ACCCTACATATCTCTGAAGCTGAAATATAGAGATGGAGTGAAGATTTAAACAGAAAT 1858  
Db 2283 ATCCCATATCTCTGAAGCTGAAATATAGAGATGAAGTGAAGATTTAAACAGAAAT 2342  
QY 1859 TAGTTACAGACATCTCAGCCAGAGAGAGAGAGAGAGAGTGAAGATTTAAACAGAGAC 1918  
Db 2343 CACTTCAGACCATCTCGGCAAGCAGAGAGAGAGAGTGAAGATTTAAACAGAGAC 2402  
QY 1919 ATGCTCCGACCAAG 1978  
Db 2403 ACCTTCACACCAAG 2462  
QY 1979 GTGAAGGCTGTGTGCTTCAAGCCCAT 2007  
Db 2463 ATGAAGGCTGTGTGCTTCAAGCCCAT 2491  
  
RESULT 3  
AB025198  
LOCUS AB025198 2176 bp mRNA linear ROD 05-AUG-2000  
DEFINITION Mus musculus mRNA for alpha-1,6-fucosyltransferase, complete cds.  
ACCESSION AB025198  
VERSION alpha-1,6-fucosyltransferase.  
KEYWORDS alpha-1,6-fucosyltransferase.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE 1 (sites)  
AUTHORS Hayashi,H., Yoneda,A. and Imamura,T.  
TITLE Molecular cloning of mouse alpha-1,6-fucosyltransferase and expression of its mRNA in the developing cerebrum  
JOURNAL DNA Seq. 11 (1-2), 91-96 (2000)  
MEDLINE 20358720  
PUBMED 10902914  
REFERENCE 2 (bases 1 to 2176)  
AUTHORS Hayashi,H., Yoneda,A. and Imamura,T.  
TITLE Direct Submission  
JOURNAL Submitted (19-MAR-1999) Toru Imamura, National Institute of Bioscience and Human Technology, Biosignaling Department; 1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan  
(E-mail:imamura@nibh.go.jp. Tel:81-298-54-6072, Fax:81-298-54-6149)  
Location/Qualifiers  
1. .2176  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"

/sex="male"		/tissue_type="cerebrum"		/dev_stage="adult"			
205..1932		/codon_start=1		/product="alpha-1,6-fucosyltransferase"			
CDS		/protein_id="BAA76392.1"		/db_xref="GI:4586554"			
		/translation="MRAWTGSRWIMLILFAWGTLTFYIGCHLVLRDNDHPDHSRELS		KILAKLERLKOQNEDELRLMAESLRIPEGPIDQGTATGRVRLVLEQLVKAKEQIENYKK			
		QARNGKDHHEILRRRENGAKELWFFLOSELKKLKHLEGNELQORHADELILLDLGHHE		RSIMTDYLYSQTGADWRKEKADITELVORRITLYLQNPQDCSKRLKLVNKNKGC			
		GYGQLHVVYCFMAYGTQRTLILESQNRWRYATGWMETVFPVSETCTDRSGLSTGH		WSGVNDKNIQVSELPIVDSLHPRPPVPLPLAPEDLADRLLRVHGDPAVMVVSQFVKY			
		LIRPQMLEKEIEBATKLGPKHPVIGVHVVRTDKVTEAFHPIIEEYMHVIEOHFQL		LARMOVDKKVELATDDPTLLKEANTKYSNYEFSIWSNGLHNRITENSIRGV			
		ILDILFLOADPLVCTFSSQVCRVAYEIMOTLHPDASANFHSLLDDIYVFGQNAHNOI		ADYPHKPRTEBEEIPMEPGDIIIGVAGNHWGYSKGINRKLKGTGLGYSYKVRKLEIVK			
		YPTYPEAK"					
BASE COUNT		669 a	474 c	494 g	539 t		
ORIGIN							
Query Match		89.9%; Score 1804.6; DB 10; Length 2176;					
Best Local Similarity		94.0%; Pred. No. 0;					
Matches 1888; Conservative		0; Mismatches 119; Indels 2; Gaps 1;					
QY	1	AACGAAACTTATTTTCTGTGTGCTAACTAGAACACAGGTACAAATGTTTCCAAATCTT	60	QY	719	CCGAAAGCTGGTATGTATATCAACAAAGGCTGTGGCTATGGATGTCAACTCCATCATG	778
DB	104	AAAGAAACCTATTTTCTGTGTGCTAACTAGAACACAGGTACAAATGTTTCCAAATCTT	163	DB	824	CCAGGAAGCTGGTGTGTAAACATCAATAAAGGCTGTGGCTATGGTGTCAACTCCATCAG	883
QY	61	TGAGCTCCGAGAAGACAGA--AGGAGTGTGAACCTCTGAAATGCGGGCATGGAATG	118	QY	779	TGGTTTACTGCTTCATGATTCCTTATGCGACCCAGCGAAACACTCATCTTGGAAATCTCAGA	838
DB	164	TGAGCTCCAGGAAGATAGAGACAGATTGAACCTCTGAAATGCGGGCATGGAATG	223	DB	884	TGGTCTACTGTTTCAAGATTCCTTATGCGACCCAGCGAAACACTCATCTTGGAAATCTCAGA	943
QY	119	CCTGCGTGTGATATGCTCATCTTTTGTGCTGGGGACCTTATGTTTATATATAGTG	178	QY	839	ATTGGCGCTATGCTACTTGGAGGATGGGAGACTGTGTTTAGACCTGTGAAGTGAAGTGA	898
DB	224	CCTGCGTGTGATATGCTCATCTTTTGTGCTGGGGACCTTATGTTTATATATAGTG	283	DB	944	ATTGGCGCTATGCTACTTGGTGGATGGGAGACTGTGTTTAGACCTGTGAAGTGAAGTGA	1003
QY	179	GTCATTTGCTGAGATATGACCACTGACCACTTCTAGCAGAGAACTCTCCAAGATTC	238	QY	899	CAGACAGGTCTGGGCTCTCCACTGGACACTGGTCAAGTGAAGTGAAGTGAAGTGAAGT	958
DB	284	GTCATTTGCTGAGATATGACCACTGACCACTTCTAGCAGAGAACTCTCCAAGATTC	343	DB	1004	CAGACAGATCTGGGCTCTCCACTGGACACTGGTCAAGTGAAGTGAAGTGAAGTGAAGT	1063
QY	239	TTGCAAGCTGGAGCGCTTAAACCAACAAATGAAGCTTGAAGGAGAAATGGCTGAGTCT	298	QY	959	AAGTGGTGGAGCTCCCATTTGTAGACAGCTTCATCTCGTCTCTTACTTACCTTTGG	1018
DB	344	TTGCAAGCTGGAGCGCTTAAACCAACAAATGAAGCTTGAAGGAGAAATGGCTGAGTCT	403	DB	1064	AAGTGGTGGAGCTCCCATTTGTAGACAGCTTCATCTCGGCTCTCTTACTTACCACTGG	1123
QY	299	TCCGAATACCAAGAGCCCTATTGATAGGGGACAGCTACAGGAAGATCCGGTGTGTTAG	358	QY	1019	CTGTACCAAGAGACTTGCAGATCGACTCTCTGAGAGTCCATGTGATCTCTGCGAGTGTGGT	1078
DB	404	TCCGAATACCAAGAGCCCTATTGATAGGGGACAGCTACAGGAAGATCCGGTGTGTTAG	463	DB	1124	CTGTTCAGAAAGACTTGCAGACCGACTCTCTTAAGAGTCCATGGTGACCCCTCGAGTGTGGT	1183
QY	359	AAGAACACTTGTAAAGCCCAAGAACACAGATTGAATAATACAGAAACAAGCTAGGAATG	418	QY	1079	GGGTATCCCAAGTGTGTCAAATACTTGTATCCGTCCACCAACCTTGGCTGGAAAGGAAATAG	1138
DB	464	AAGAACACTTGTGTAAAGCCCAAGAACACAGATTGAATAATACAGAAACAAGCTAGGAATG	523	DB	1184	GGGTGTCCCAAGTGTGTCAAATACTTGTATCCGTCCACCAACCTTGGCTGGAAAGGAAATAG	1243
QY	419	ATCTGGGAAAGGATCATGAATCTTAAAGGAGGAGATTGAAATGGAGCTTAAAGAGCTCT	478	QY	1139	AAGAAACCAACAAAGTGTGCTTCAAAATCCAGTATTTGGAGTCCATGTCCAGACGCA	1198
DB	524	GTCTGGGAAAGGATCATGAATCTTAAAGGAGGAGATTGAAATGGAGCTTAAAGAGCTCT	583	DB	1244	AAGAAACCAACAAAGTGTGCTTCAAAATCCAGTATTTGGAGTCCATGTCCAGACGCA	1303
QY	479	GGTTTTTCTCAAAAGTGAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAAT	538	QY	1199	CTGACAAAGTGGGACAGAGACAGCTTCCATCCATTTGAGGAATACATGTGACAGCGCA	1258
DB	584	GGTTTTTCTCAAAAGTGAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAAT	643	DB	1304	CAGACAAAGTGGGACAGAGACAGCTTCCACCCCTCGAGAGTACATGTGACAGCGTTG	1363
QY	539	GACATGACAGTGAATTTCTTTTGGATTTAGGACATCATGAAGTCTTATCATGACAGATC	598	QY	1259	AAGAAACATTTTCAAGTCTCTGAAACGCGAGAAATGAAGTGGATGAAGAAAGAGTGTATCTGG	1318
DB	644	GACATGACAGTGAATTTCTTTTGGATTTAGGACATCATGAAGTCTTATCATGACAGATC	703	DB	1364	AACAAACATTTTCAAGTCTCTGCAAGCAGAGATGCAAGTGGATGAAGAAAGAGTATATCTGG	1423
QY	599	TATACCTACTCAGTCAAAACAGATGAGCAGGTGAGTGGCGGAAAGAAAGAAAGAAAGATC	658	QY	1319	CCACTGATGACCTTCTTTGTTAAAGGAGGCAAGACAAAGTACTCCAAATATGAATTTA	1378
DB	704	TATACCTACTCAGTCAAAACAGATGAGCAGGTGAGTGGCGTGAAGAAAGAGCCAAAGATC	763	DB	1424	CTACTGATGATCTCTACTTTGTTAAAGGAGGCAACACAAAGTACTCCAAATATGAATTTA	1483
QY	659	TGACAGAGCTGTCCAGCGGAGAAATACATATCTGCAAGATCCCAAGGACTGCAGCAAG	718	QY	1379	TTAGTGAATCTCTATTTCTTTGTCAGCTGAGCTACACACCGATACACAGAAATTCAC	1438
DB	764	TGACAGAGCTGTCCAGCGGAGAAATACATATCTGCAAGATCCCAAGGACTGCAGCAAG	823	DB	1484	TTAGTGAATCTCTATTTCTTTGTCAGCTGAGCTACACAAATCGGTACACAGAAATTCAC	1543

```
QY 1799 ACCCTACATATCTCTGAAGCTGAAATATAGAGATGAGATGAGATTAAGAGATTAAACAAGAAATT 1858
Db 1904 ATCCACATATCTCTGAAGCTGAAATATAGAGATGAGATGAGATTAAGAGATTAAACAAGAAACT 1963
QY 1859 TAGTTGACACCATCTCAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1918
Db 1964 CACTTCAGACCATCTCTGGCCCAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2023
QY 1919 ATGCTCCGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1978
Db 2024 AGCTTCCACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2083
QY 1979 GTGAAGGCTGCTGTGCCCTCAAGCCCAT 2007
Db 2084 ATGAAGGCTGTGGTGCCCTCAAGCCCAT 2112

RESULT 4
LOCUS BD168517 1728 bp DNA linear PAT 17-JAN-2003
DEFINITION Cells producing antibody composition.
ACCESSION BD168517
VERSION BD168517.1 GI:27874329
KEYWORDS WO 0231140-A/2.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1728)
AUTHORS Kanda, Y., Sato, M., Nakamura, K., Uchida, K., Shinkawa, T., Yamane, N.,
Hosaka, E., Yamasaki, M. and Hanai, N.
TITLE Cells producing antibody composition
JOURNAL Patent: WO 0231140-A 2 18-APR-2002;
KYOWA HAKKO KOGYO CO LTD
COMMENT OS Mus musculus (mouse)
PN WO 0231140-A/2
PD 18-APR-2002
PF 05-OCT-2001 WO 2001JP008804
PR 06-OCT-2000 JP 00P 308526
PI YUTAKA KANDA, MITSUO SATO, KAZUYASU NAKAMURA, KAZUHIISA UCHIDA, PI
TOYOHIDE SHINKAWA, NAOKO YAMANE, EMI HOSAKA, MOTOO YAMASAKI, NOBUO
PI HANAI
PC C12N5/10, C12P21/08, C07K16/00, A01K67/00, A61K39/395, C12N9/00, PC
C12N15/52,
PC G01N33/53
CC Cells producing antibody composition
FH Key Location/Qualifiers
FT source 1. .1728
FT Location/Qualifiers
source 1. .1728
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"

BASE COUNT 529 a 365 c 400 g 434 t
ORIGIN

Query Match 78.9%; Score 1584; DB 6; Length 1728;
Best Local Similarity 94.8%; Pred. No. 0;
Matches 1638; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 100 ATCGGGGATGGACTGGTTCCTGGCGTTGGATATGCTCATTTTTCCTGGGGGACC 159
Db 1 ATCGGGGATGGACTGGTTCCTGGCGTTGGATATGCTCATTTTTCCTGGGGGACC 60

QY 160 TTATTTTATATAGTGGTTCATTTGGTTCAGATAATGACCCCTGACCATTTCTAGC 219
Db 61 TTGTTATTTATATAGTGGTTCATTTGGTTCAGATAATGACCCCTGACCATTTCTAGC 120

QY 220 AGAGAACTCTCAAGATTTCTGCAAGCTGGAGCGCTTTAAACACACAAATGAAGACTTG 279
Db 121 AGAGAACTCTCAAGATTTCTGCAAGCTTTGAACGCTTTAAACACACAAATGAAGACTTG 180
```

```
QY 280 AGGAAGATGGCTGAGTCTCTCCGAATACCAAGGCCCTATTGATCAGGGGACAGCTACA 339
Db 181 AGGCGAATGGCTGAGTCTCTCCGAATACCAAGGCCCTATTGACCAAGGGGACAGCTACA 240
QY 340 GGAAGAGTCGCTGTTTGTAGGAAGAACAGCTTGTTAAGGCCAAAGAACAGATTTGAANAATTAC 399
Db 241 GGAAGAGTCGCTGTTTGTAGGAAGAACAGCTTGTTAAGGCCAAAGAACAGATTTGAANAATTAC 300
QY 400 AAGAAAACAAGCTAGGAATGATCTGGGAAAGGATCATGAAATCTTTAAGGAGGAGGATTTAA 459
Db 301 AAGAAAACAAGCTAGGAATGATCTGGGAAAGGATCATGAAATCTTTAAGGAGGAGGATTTAA 360
QY 460 AATGGAGCTAAAGAGCTCTGGTTTTTTTCTACAAAGTGAATTTGAAGAAATTTAAAGAAATTA 519
Db 361 AATGGAGCTAAAGAGCTCTGGTTTTTTTCTACAAAGTGAATTTGAAGAAATTTAAAGAAATTA 420
QY 520 GAAGGAACGAATCTCCAAAGACATGACAGATGAAATCTTTTGGATTTTAGGACATCATGAA 579
Db 421 GAAGGAATGAATCTCCAAAGACATGACAGATGAAATCTTTTGGATTTTAGGACATCATGAA 480
QY 580 AGGTCTATCATGACAGATCTATCTACTCAGTCAAAACAGATGGAGCAGGTGAGTGGCGG 639
Db 481 AGGTCTATCATGACAGATCTATCTACTCAGTCAAAACAGATGGAGCAGGTGAGTGGCGG 540
QY 640 GAAAAAGAACCAAGATCTGACAGAGCTGGTCAGCGGAGAAATAACATATCTCCAGAA 699
Db 541 GAAAAAGAGGCCAAAGATCTGACAGAGCTGGTCAGCGGAGAAATAACATATCTCCAGAA 600
QY 700 CCCAAGGACCTGCACGAAGCCAGAAAGCTGGTATGTAATCAACAAGGCTGGGCTAT 759
Db 601 CCTAAGGACCTGCACGAAGCCAGAAAGCTGGTATGTAATCAACAAGGCTGGGCTAT 660
QY 760 GGATGTCAACTCCATCATGTGGTTTTACTGTCTTATGATTTGCTTATGGCACCACCGCAACA 819
Db 661 GGTGTCAACTCCATCATGTGGTTTTACTGTCTTATGATTTGCTTATGGCACCACCGCAACA 720
QY 820 CTCAATCTTGAATCTCAGAAATTTGGCGCTATGCTACTGAGAGATGGGAGAGCTGTGTTAGA 879
Db 721 CTCAATCTTGAATCTCAGAAATTTGGCGCTATGCTACTGAGAGATGGGAGAGCTGTGTTAGA 780
QY 880 CCTGTAAGTGAAGATCCACAGACAGCTCTGGCGCTCTCCACTGGACACTGGTCAGGTGAA 939
Db 781 CCTGTAAGTGAAGATCCACAGACAGCTCTGGCGCTCTCCACTGGACACTGGTCAGGTGAA 840
QY 940 GTGAAGGACAAAAATGTTCAAGTGGTCAGCTCCCAATTTGTAGACAGCTCCATCTCGT 999
Db 841 GTAAATGACAAAAACATTCAGTGGTCAGCTCCCAATTTGTAGACAGCTCCATCTCGG 900
QY 1000 CTTCTTTACTTACCTTGGCTGTACCAAGAGACCTTGCAGATCGACTCTCGAGAGTCCAT 1059
Db 901 CTTCTTTACTTACCTTGGCTGTTCAGAAAGACCTTGCAGACCGACTCTTAAGAGTCCAT 960
QY 1060 GGTGATCTCAGTGTGGTATCCAGTTTGTCAAATACTTTGATCCGTCACCAACCT 1119
Db 961 GGTGATCTCAGTGTGGTATCCAGTTTGTCAAATACTTTGATTCGTTCCCAACCT 1020
QY 1120 TGGCTGGAAGGGAATAGAGAAACCAAGAGCTTTGGCTTTCAAAACATCCAGTTATT 1179
Db 1021 TGGCTGGAAGGGAATAGAGAAAGCCACCAAGAGCTTTGGCTTTCAAAACATCCAGTTATT 1080
QY 1180 GGAGTCCATGTTCAGACCGCATGACAAAGTGGGAAACAGAGAGCCTTCCATCCCATTCGAG 1239
Db 1081 GGAGTCCATGTTCAGACCGCATGACAAAGTGGGAAACAGAGAGCCTTCCATCCCATTCGAG 1140
QY 1240 GAATACATGTTACAGTTTGAAGAACATTTTCAGCTTCTCGAACCGCAGAAATGAAGTGGAT 1299
Db 1141 GAGTACATGTTACAGTTTGAAGAACATTTTCAGCTTCTCGAACCGCAGAAATGAAGTGGAT 1200
QY 1300 AAAAAAGAGTGTATCTGGCCACTGATGACCCCTTTTGTGTTAAAGAGGCAAGACAAAG 1359
Db 1201 AAAAAAGAGTGTATCTGGCTACTGATGATCTTACTTGTGTTAAAGAGGCAAGACAAAG 1260
QY 1360 TACTCCAAATTATGAATTTATTAGTGATACTTATTTCTTGGTTCAGCTGAGCTACACAA 1419
```



```

Db      1261 TACTCCAAATATGAATTTATTAGTGATAACTCTATTTCTTGCTCAGCTGGACTACAAAT 1320
Qy      1420 CGATACACAGAAATTCATCTCGGGCGTGATCCTGGATATACACTTCTCTCCAGGCT 1479
Db      1321 CGGTACACAGAAATTCATCTCGGGCGTGATCCTGGATATACACTTCTCTCCAGGCT 1380
Qy      1480 GACTTCTCTGTGTGTACTTTTTTCATCCAGGCTCTGTAGGGTGTCTTATGAATCATGCAA 1539
Db      1381 GACTTCTCTGTGTGTACTTTTTTCATCCAGGCTCTGTAGGGTGTCTTATGAATCATGCAA 1440
Qy      1540 ACATGCAATCTGTAGTCTCTGCAAACTTCATCTCTTTAGATGACATCTACTATTTTGA 1599
Db      1441 ACCCTGCAATCTGTAGTCTCTGCAAACTTCATCTCTTTAGATGACATCTACTATTTTGA 1500
Qy      1600 GGCCAAATGCCCAACACAGATTCGAGTTTATCTCACCACCTCGAACTCGAACTGAAGAGAA 1659
Db      1501 GGCCAAATGCCCAACATCAGATTCGAGTTTATCTCACCACCTCGAACTCGAACTGAAGAGAA 1560
Qy      1660 ATCCCATGGAACTCGAGATATCATTTGCTGTGGCTGGAAACCATTTGAATGGTTACTCT 1719
Db      1561 ATCCCATGGAACTCGAGATATCATTTGCTGTGGCTGGAAACCATTTGGATGGTTATCT 1620
Qy      1720 AAAGGTGCAACAGAAACTAGGAAACACAGGCTGTACCTTCTTACAAAGTCCGAGAG 1779
Db      1621 AAAGGTATCAACAGAAACTTGGAAACACAGGCTTATATCCTCTCTCAAAAGTCCGAGAG 1680
Qy      1780 AAGATAGAAACAGTCAATACCTACATATCTCTGAGCTGAAAAATAG 1827
Db      1681 AAGATAGAAACAGTCAAGTATCCCATATCTCTGAGCTGAAAAATAG 1728

```

## RESULT 5

HSAS39535

LOCUS

DEFINITION Homo sapiens mRNA for alpha6-fucosyltransferase (FUT8 gene), clone CD-L1.

ACCESSION

AJ539535

VERSION

AJ539535.1

KEYWORDS

alpha6-fucosyltransferase; FUT8 gene.

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

Mollicone, R., Michalski, J.C., Bauvy, C., Cailliau-Thomas, A.,

Candelier, J.J., Martinez-Duncker, I., Breton, C., Codogno, P. and

Oriol, R.

TITLE

Splice variants of alpha6-fucosyltransferase are expressed early in

human embryogenesis

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 2796)

AUTHORS

Oriol, R.

TITLE

Direct Submission

JOURNAL

Submitted (31-JAN-2003)

Vaillant-Coururier, 94807, FRANCE

FEATURES

Location/Qualifiers

1..2796

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/chromosome="14"

/map="14q23.2"

/clone="CD-L1"

/tissue\_type="whole embryo"

/dev\_stage="40-70 days"

/country="France:Paris"

1..2796

/gene="FUT8"

326..2053

/gene="FUT8"

/EC\_number="2.4.1.68"

/function="adds fucose in alpha1,6 linkage to the internal

GlcNAc of the core chitobiose"

/codon\_start=1

/evidence=experimental

/product="alpha6-fucosyltransferase"

/protein\_id="CAD62444.1"

/db\_xref="GI:28200808"

/translation="MRPWTGSRWIMLILFAWGTLLFYIGHILVRDNDHPDHSRELS

KILAKLERLQKQNEDELRRMAESLRIPGPIDQGPAGIRVRLBEQLVKAKEQIENYKK

RSIMTDLXYLSOTDAGDWEKAKOLTELVRRIITYLONPKDCSKAKLVNINKGC

GYCOLHIVYCFMIAIYQRTILILSQNRWYATGWETVPRVSECTDERSGISTGH

WSGEVDKNNQVVELPILVDSLRPPPLPLAVPEDLADRLVRVRDPAVWVVSQFVK

LIRPQWLEKEIBEATKGLGHEVGVHVRRTDKVGTAAFPPIESIMVWVSHFOL

LARMQVKKRVYLATDPPSLKKEIKYNYEFSIDNSISWAGLHNRITYENSLRGV

ILDIHLSQADFLVCTPSSQVCRVAYEMOTLHPDASANFHSLLDDIYFGQNAHNOI

AIYAHOPFTADEIPMBEGDIIIGVAGNHWGYSKGVNRKLGRTGLYPSYKVKREKLETYK

YPTYPEAK"

2766..2771

/gene="FUT8"

polyA\_signal

polyA\_site

BASE COUNT 874 a 552 c 605 g 765 t

ORIGIN

Query Match 78.4%; Score 1575.2; DB 9; Length 2796;

Best Local Similarity 87.5%; Pred. No. 0;

Matches 1781; Conservative 0; Mismatches 223; Indels 32; Gaps 4;

Qy 1 AACGAAACTTATTTCTGTGTGGCTAACTAGAACCCAGAGTACAAATGTTTCCAAATCTT 60

Db 225 AACGAAAGTCTATTTCACTGTGCACTAACTAGAACCCAGAGTACAAATGTTTCCAAATCTT 284

Qy 61 TGAGCTCCGAGAGACACA--GAAGGAGTGTGAACTCTGAAATCGCGGCATGGACTGTT 118

Db 285 TGAGCTCCAGAGCTCCAGGGAAGTGTGAAATCTGAAATCGCGGCATGGACTGTT 344

Qy 119 CTGCGCTGTGGATTATGCTCATTTCTTTGCTGGGGACCTTTATTTTATATAGTG 178

Db 345 CTGCGCTGTGGATTATGCTCATTTCTTTGCTGGGGACCTTTGTTTATATAGTG 404

Qy 179 GTCAATTTGTTTGGAGTAATGACCCCTGACCATTTCTAGCAGAGAACTCTCCAGATTC 238

Db 405 GTCACTTTGGTACGAGATAATGACCATCTCTATCATCTAGCCGAGAACTGTCCAGATTC 464

Qy 239 TTGCAAGCTGGAGCGCTTAAACCAACAAATGAAGACTTTGAGGAGAAATGCTCAGTCTC 298

Db 465 TGGCAAGCTTGAAGCTTAAACCAACGAGAAATGAAGACTTTGAGGCGGAATGCCGAATCTC 524

Qy 299 TCCGAATACCCAGAGGCTCTTATTTGATCAGGGGACAGCTACAGGAAGAGTCCGTTTGTAG 358

Db 525 TCCGATACCCAGAGGACCTTATTTGATCAGGGGACAGCTATAGGAAGAGTACGCGTTTGTAG 584

Qy 359 AAGAACAGCTTTTAAAGCCAAAGAAACAGATTGAAATTTACAGAAACAGCTAGGAATG 418

Db 585 AAGAACAGCTTTTAAAGCCAAAGAAACAGATTGAAATTTACAGAAACAGCTAGGAATG 644

Qy 419 ATCTGGGAAAGGATCATGAAATCTTAAAGGAGGAGATTGAAATTTGGAGCTTAAAGAGCTCT 478

Db 645 GTCTGGGAAAGGATCATGAAATCTTAAAGGAGGAGATTGAAATTTGGAGCTTAAAGAGCTCT 704

Qy 479 GGTTTTTTCTACAAAGTGAATTTAAAGAAATTTAAAGAAATTTAGAAAGAAACCTCCAAA 538

Db 705 GGTTTTTTCTACAGATGAATTTAAAGAAATTTAAAGAAATTTAGAAAGAAATTTAGAAATTTCCAAA 764

Qy 539 GACATGAGATGAATTTTGTGATTTAGGACATCATGAAAGCTCTATCATGACAGATC 598

Db 765 GACATGAGATGAATTTTGTGATTTAGGACATCATGAAAGCTCTATCATGACAGATC 824

Qy 599 TATACCTACCTCAGTCAAAACAGATGAGGAGGAGTGGCGGGAAGAAAGCCAAAGATC 658

Db 825 TATACCTACCTCAGTCAAAACAGATGAGGAGGAGTGGCGGGAAGAAAGCCAAAGATC 884

Qy 659 TGACAGAGCTGGTCCAGCGGAGAAATACATATCTGCAGAAATCCCAAGGAGCTGCAGCAAG 718





KILAKLERLKOQNEBRLMRBLSRIPEPIDQGPALGRVRLBQVLKAKEQIENYKK  
RTNGLGKHDSILERRINGAKELWFFLOSELKKLKNLEGNELQORHAEFLDLIGHHE  
RSIMDLVSLQAGDWREKEADLTTELVRITLYLONPKDCSKAKKLVNINKGC  
GYCQLHHVYCPMAYGTQRTLLILESONWRYATGGWETVFPVSECTDRSGISTGH  
WSGEVKQNVQVSELPIVDSUHPRPYLPPLAVPEDLADRLVRVHGDPAWVWSQFVKY  
LIRPQWLEKEIEBATEKLGKFKPVIQVHVRRTDKVGTPEAAFPPIBEYMHVESHFOL  
LARMOVDKKEVYLATDDPSLLKEAKTKYPNYEFISONSISWSAGLHNRVYENSLSRGV  
ILDHFLSOADFLVCTSSQVCRVAYEIMQTLHPDASANFHSLODIYVFGQNAHNOI  
AIYAHOPRTADEIPMEPGDIIIGVAGNHWDGYSKGVNRKLGRTGLYPSYKVRKEIETVK  
YPTYPEABK"

## variation

2218

/gene="FUT8"

/note="II"

/frequency=".5"

/replace="c"

2758..2763

/gene="FUT8"

/note="putative"

2770..2775

/gene="FUT8"

/note="putative"

2869..2874

/gene="FUT8"

/note="putative"

BASE COUNT

881 a 589 c 641 g 787 t

ORIGIN

Query Match 78.4%; Score 1575.2; DB 9; Length 2898;

Best Local Similarity 87.5%; Pred. No. 0;

Matches 1781; Conservative 0; Mismatches 223; Indels 32; Gaps 4;

QY 1 AACGAACTTATTTTCGTGTGCTAACTAGAACACGAGGTACAACTGTTTCCAACTCTT 60  
DB 328 AACAGAGCTTATTCACCTGTGCTACTAATAGAAACAGAGTTACAAAGTTTCAATCTT 387  
QY 61 TGAGCTCCGAGAAGACA--GAAGGAGTTGAAACTCTGAAAATCGGGCATCGACTGGTT 118  
DB 398 TGAGCTCCGAGACTCCAGGGAAGTGAGTTGAAAATCTGAAAATCGGGCATCGACTGGTT 447  
QY 119 CTGCGGTGGATGATGCTCACTCTTTTGGCTGGGGACCTTATGTTTATATAGTG 178  
DB 448 CTGCGGTGGATGATGCTCACTCTTTTGGCTGGGGACCTTATGTTTATATAGTG 507  
QY 179 GTCACTTCTGTCAGATAATGACACCTGACCACTTCTAGCAGAGAACTCTCCAGATTC 238  
DB 508 GTCACTTCTGTCAGATAATGACCACTCTGATCACTCTAGCCGAGAACTCTCCAGATTC 567  
QY 239 TTGCAAACTGGAGCGCTTAAACAAACAAAATGAAGACTTGAGGAGAAATGGCTGAGTCTC 298  
DB 568 TGCAAACTTGAACGCTTAAACAGCAGAAATGAAGACTTGAGGCGAATGGCCGAATCTC 627  
QY 299 TCCGAATACGAGAGCGCTTATGATAGGGGACAGCTACAGGAAGAGTCCGTGTTTATG 358  
DB 628 TCCGATACGAGAGCGCTTATGATAGGGGCGCAGCTATAGGAAGAGTACGCGCTTTAG 687  
QY 359 AAGACACTCTTAAAGCCAAAGAACAGATTGAAAATTAAGAAACAAAGCTAGGAATG 418  
DB 688 AAGACACTCTTAAAGCCAAAGAACAGATTGAAAATTAAGAAACAAAGCTAGGAATG 747  
QY 419 ATCTGGGAAAGGATCATGAAATCTTAAAGGAGAGGATTCGAAATGGAGCTTAAAGAGCTCT 478  
DB 748 GTCTGGGAGGATCATGAAATCTTAAAGGAGAGGATTCGAAATGGAGCTTAAAGAGCTCT 807  
QY 479 GGTCTTTTCTCAAAAGTGAATTAAGAAATTAAGAAATTAAGAAACAAAGCTAGGAATG 538  
DB 808 GGTCTTTTCTCAAAAGTGAATTAAGAAATTAAGAAATTAAGAAACAAAGCTAGGAATG 867  
QY 539 GACATGACAGTCAAAATCTTTTGGATTTAGGACATCATGAAAGTCTATCATGACAGATC 598  
DB 868 GACATGACAGTCAAAATCTTTTGGATTTAGGACATCATGAAAGTCTATATATGCGGATC 927  
QY 599 TATACCTACCTAGTCAAAACAGATGAGCAGGTGAGTGGCGGGAAGAAAGCAAGATC 658  
DB

DB 928 TATACTACTCAGTCAGTCAGACAGATGGAGCAGGTGATTGGCGGAAAAAGAGCCAAAGATC 987  
QY 659 TGACAGAGCTGGTCCACGCGAGATAACATATCTGCAGAAATCCCAAGGACTGCAGCAAG 718  
DB 988 TGACAGAACTGGTTCACGCGAGATAACATATCTTCAGATCCCAAGGACTGCAGCAAG 1047  
QY 719 CCAGAAAGCTGGTATTAATATCAAAAGGCTGTGGCTATGAGTGTCAATCTCCATCATG 778  
DB 1048 CCAAAAAGCTGGTGTGTAATATCAAAAGGCTGTGGCTATGAGTGTGAGTGTCCATCATG 1107  
QY 779 TGGTTACTGCTTCAGTATGCTTATGGCAGCCAGGAAACACTCATCTTGAATCTCAGA 838  
DB 1108 TGGTCTACTGCTTCAGTATGCTATGGCAGCCAGGCAACACTCATCTTGAATCTCAGA 1167  
QY 839 ATTGGCGCTATGCTACTTGGAGGATGGGAGACTGTGTTTAGACCTGTAGTGAACATGCA 898  
DB 1168 ATTGGCGCTATGCTACTTGGTGGATGGGAGACTGTATTTAGGCTGTAGTGAACATGCA 1227  
QY 899 CAGACAGGTCTGGCTCTCCACTGGACACTGGTCAAGTGAAGTGAAGGACAAAAATGTTTC 958  
DB 1228 CAGACAGATCTGGCATCTCCACTGGACACTGGTCAAGTGAAGTGAAGGACAAAAATGTTTC 1287  
QY 959 AAGTGGTGGAGCTCCCATTTGTAGACAGCTCCATCTCTGCTCTCTTACTTACCTTGG 1018  
DB 1288 AAGTGGTGGAGCTCCCATTTGTAGACAGCTCTTATCCCGCTCTCCATATTTACCTTGG 1347  
QY 1019 CTGTACACAGAGACTTGGCAGATCGACTCTCTGAGAGTCCATGAGTGTCTCTCAGTGTGCT 1078  
DB 1348 CTGTACACAGAGACTTGGCAGATCGACTCTGAGAGTGTGAGTGTGAGTGTGAGTGTGCT 1407  
QY 1079 GGGTATCCCAATTTGTCTCAAAATCTTGTATCCCTCCACAACTTGGCTGTGAAAAGGAAATPAG 1138  
DB 1408 GGGTGTCTCAGTTTGTCAAAATCTTGTATCCCGCTCCACAGCTTGGCTAGAAAAGAAATPAG 1467  
QY 1139 AAGAAAACACCAAGAACTTGGCTTCAAACTCAGATTTATGAGTGTGAGTGTGAGTGTGAGTGTG 1198  
DB 1468 AAGAAAGCAACCAAGAACTTGGCTTCAAACTCAGATTTATGAGTGTGAGTGTGAGTGTGAGTGTG 1527  
QY 1199 CTGACAAAGTGGGAAAGAGCAGCTTCCATCCATTTGAGGAAATACATGTCAGACGTTG 1258  
DB 1528 CAGACAAAGTGGGAAAGAGCAGCTTCCATCCATTTGAGGAAATACATGTCAGTGTGATGTTG 1587  
QY 1259 AAGAACTATTTTCACTCTTCTGAAACGCAAGATGAAAGTGGATTAAGAAAGAGTGTATCTGG 1318  
DB 1588 AAGAACTATTTTCACTCTTCTGAAACGCAAGATGAAAGTGGATTAAGAAAGAGTGTATGTTGG 1647  
QY 1319 CCATGATGACCTCTTCTTAAAGGAGGCAAGACAAAGTACTTCCAAATTTATGAATTTA 1378  
DB 1648 CCACAGATGACCTCTTCTTAAAGGAGGCAAGACAAAGTACTTCCAAATTTATGAATTTA 1707  
QY 1379 TTAGTGTAACTCTATTTCTTGGTTCAGTGGACTTACACACCGATACACAGAAATTTAC 1438  
DB 1708 TTAGTGTAACTCTATTTCTTGGTTCAGTGGACTTACACACCGATACACAGAAATTTAC 1767  
QY 1439 TTCCGGCGCTGATCCTGATATACCTTCTCTCCAGGCTGACTTCTCTGTGTGTACTT 1498  
DB 1768 TTCTGTGAGTGTATCTGATATATATTTCTCTCTCAGGACAGCTTCTCTAGTGTGTACTT 1827  
QY 1499 TTTTCATCCAGGCTGTAGGGTGTCTTATGAATTCATGCAAAACCTGATCCTGATGCT 1558  
DB 1828 TTTTCATCCAGGCTGTAGGGTGTCTTATGAATTCATGCAAAACCTGATCCTGATGCT 1887  
QY 1559 CTGCAAACTTCCATTTCTTAGATGACATCTACTATTTTGGAGGCGCAAAATGCCACACACC 1618  
DB 1888 CTGCAAACTTCCATTTCTTAGATGACATCTACTATTTTGGAGGCGCAAAATGCCACACACC 1947  
QY 1619 AGATTGCAAGTTTATCTCTTCAACCACTCGAACTAAAGAGGAAATCCCATGGAACCTGGAG 1678  
DB 1948 AAATGCAATTTATGCTTACCAACCCGAACTCGAGTGAATTTCCCATGGAACCTGGAG 2007  
QY 1679 ATATCATTTGGTGTGGCTGGAACCAATTTGGAATGGTGTACTCTTAAAGGTGTCAACAGAAAC 1738  
DB 2008 ATATCATTTGGTGTGGCTGGAATCATTTGGGATGCTTATTTAAAGGTGTCAACAGGAAAT 2067



Qy 839 ATTGGCGTACTGCTACTGGAGGATGGGAGACTGTGTTTATAGACCTGTGTAAGTGAGACATGCA 898  
Db 1065 ATTGGCGTACTGCTACTGGTGGTGGAGACTGTATTTAGCGCTGTAAAGTGAGACATGCA 1124  
Qy 899 CAGACAGGTCTGGCTCTTCCACTGCACTCGACTCGGTGAGGTGAAGTGAAGGACAAAATGTTT 958  
Db 1125 CAGACAGATCTGGCATCTCCACTGCACTCGACTCGGTGAGGTGAAGTGAAGGACAAAATGTTT 1184  
Qy 959 AAGTGGTCTGAGCTCCCATTTGTAGACAGCTCCATCTCGTCCCTTACTTACCCCTGG 1018  
Db 1185 AAGTGGTCTGAGCTCCCATTTGTAGACAGCTCTTCATCCCGCTCCATATTTACCCCTGG 1244  
Qy 1019 CTGTACCAAGAAGACTTTCAGATCCACTCTGAGAGTCCATGGTGCATCTCGCTGAGTGGT 1078  
Db 1245 CTGTACCAAGAAGACTTTCAGATCCACTCTGAGAGTCCATGGTGCATCTCGCTGAGTGGT 1304  
Qy 1079 GGGTATCCAGTTTGTCAAACTTGTATCTGATCTCCACAACTTGGCTGGAAGGAAATAG 1138  
Db 1305 GGGTGTCTCAGTTGTCAAACTTGTATCTGATCTCCCGCCACAGCTTGGCTAGAAAAGAAATAG 1364  
Qy 1139 AAGAAACCAAGAGCTGGCTTCAAACTTGTATCTGATCTCCAGTTCATGAGTCCATGTCAGACGCA 1198  
Db 1365 AAGAAACCAAGAGCTGGCTTCAAACTTGTATCTGAGTTCATGAGTCCATGTCAGACGCA 1424  
Qy 1199 CTGCAAAAGTGGGAACAGAGCTTCCATCTCCATTCAGGAAATACATGTTGACACGTTG 1258  
Db 1425 CAGCAAAAGTGGGAACAGAGCTTCCATCTCCATTCAGGAAATACATGTTGACACGTTG 1484  
Qy 1259 AAGAACTTTTTCAGTCTTCGAGCGCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1318  
Db 1485 AAGAACTTTTTCAGTCTTCGAGCGCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1544  
Qy 1319 CCATGATGACCTCTCTTTTAAAGGAGGCAAGAGTACTTCCAAATTTATGAATTTA 1378  
Db 1545 CCACAGATGACCTCTCTTTTAAAGGAGGCAAGAGTACTTCCAAATTTATGAATTTA 1604  
Qy 1379 TTAGTGAATCTCTATTTCTTGGTCTGAGTCTGACATACAACTGACATACAGAAATTCAC 1438  
Db 1605 TTAGTGAATCTCTATTTCTTGGTCTGAGTCTGACATACAACTGACATACAGAAATTCAC 1664  
Qy 1439 TTGGGGGCGTATCTCGATATACATTTCTCTCCAGGCTGACTTCTTGTGTGACTT 1498  
Db 1665 TTGGTGGAGTATCTCGATATACATTTCTCTCTCAGCGAGACTTCTCTAGTGTGACTT 1724  
Qy 1499 TTTCATCCAGTCTGTAGGTTGCTTATGAATCATGCAACACTGATCTGATGCT 1558  
Db 1725 TTTCATCCAGTCTGTAGGTTGCTTATGAATCATGCAACACTGATCTGATGCT 1784  
Qy 1559 CTGCAAACTTCCATTTTATGATGATCATCTACTATTTTGGAGGCCAATAATGCCACAC 1618  
Db 1785 CTGCAAACTTCCATTTTATGATGATCATCTACTATTTTGGAGGCCAATAATGCCACATC 1844  
Qy 1619 AGATTGCAAGTTTATCTCCACCACTCTGAACTTAAAGAGGAAATCCCAATGAACTGGAG 1678  
Db 1845 AAATTGCCATTTATCTCACCACCCCGAATGCAAGTGAATTTCCCATGCACTGGAG 1904  
Qy 1679 ATATCATTTGGTGGCTGGAACCACTTGAATGGTGTACTCTAAAGGTCTCAACAGAAAC 1738  
Db 1905 ATATCATTTGGTGGCTGGAATCATTTGGATGGCTTATTTAAAGGTGTCAACAGAAAT 1964  
Qy 1739 TAGGAAACACAGGCTGTACCTTCTCAAGTCCGAGAGAGATGAACACAGTCAAT 1798  
Db 1965 TGGGAAGCAGGCGCTATATCCCTCTCAAGTTCGAGAGAGATGAACACGCTCAAGT 2024  
Qy 1799 ACCCTACATATCTGAAGCTGAAATATAGAGTGGAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1848  
Db 2025 ACCCAATATCTGAGGCTGGAATTAAGTCTAGATGGAAGATTAACCAACCAACT 2084  
Qy 1849 -----CAACAGAAATTTAGTTTACAGACCTCTCAGCCCAAGCAGAACCCAG 1893  
Db 2085 CAGTTTCGACCAACTCAGTTCAACACCTTTGAGCCAACTGTAGATGAAGAGGGCTCTGA 2144

1894 ----ACTAACATATGTTTATGACAGACATGCTCCGACCAAGAGCAAGTGGGAACCT 1949  
2145 TCTAACAAAATAAGTTATATAGTAGTACTCTCTCAGCACCAGAGCAGCTGGAACTGA 2204  
Qy 1950 CAGATCTGTC-ACTGTGGGAACCTCTTTGTGAAGGCTGCTGTGCTCCCTCAAGCC 2004  
Db 2205 CATAGGCTTCAATGTGGGAATTCCTCTTTTAAAGGGCTGCAATGCTCATATCC 2260

RESULT 8  
HSY17977  
LOCUS HSY17977 2992 bp mRNA linear PRI 17-SEP-1998  
DEFINITION Homo sapiens mRNA for glycoprotein 6-alpha-L-fucosyltransferase,  
transcript B2.  
ACCESSION Y17977  
VERSION Y17977.1 GI:3451264  
KEYWORDS FUT8 gene; glycoprotein 6-alpha-L-fucosyltransferase.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1  
AUTHORS Cailliau,A., Balanzino,L., Candelier,J.J., Oriol,R. and  
Mollicon,R.  
TITLE Differential splice variants of human FUT8 embryonic cDNA  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2992)  
AUTHORS Cailliau,A.L.V.  
TITLE Direct Submission  
JOURNAL Submitted (19-AUG-1998) A.L.V. Cailliau, INSERM U178, 16 av P.V.  
Couturier, 94807 Villejuif Cedex, FRANCE  
COMMENT Related entry: D89289.  
FEATURES  
1..2992  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/isolate="transcript B2"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/map="14q33"  
/tissue\_type="whole embryo"  
/dev\_stage="40-70 days"  
1..2992  
/gene="FUT8"  
1..101  
/gene="FUT8"  
/note="B"  
102..200  
/gene="FUT8"  
/note="C"  
201..2156  
/gene="FUT8"  
/note="D"  
429..2156  
/gene="FUT8"  
/EC\_number="2.4.1.68"  
/note="allele 1"  
/codon\_start=1  
/product="glycoprotein 6-alpha-L-fucosyltransferase"  
/protein\_id="CAA76986.1"  
/db\_xref="GI:3451265"  
/db\_xref="SPTREMBL:O00235"  
/translation="MRPTGSRWMLILFAWGTLLFYIGGLVLRDNDHPDSSRELS  
KILAKRLKQONDLRRMAESLRIPGPIDQGAIGRVRLERQLVAKAQENYKK  
RTRNGDKHEILRRRIENGAKLWFFLSELKKLKNLEGNLQRHADFLDLGHHH  
QSNIMDLVYLQSDTGAGDREKEADLTQLVRRITLYLQNPDCSKAKKLVNKKGC  
GYGQLHHVYCFMAYQRTLLLESQNRWYATGWTVPFVPTSETCTDRSGISTGH  
WGSVKDKNVQVVELPIVDSLHPRPVPLAVPELDRLVRVHGDPAVMWVSPVKY  
LIRQPMLEKIEEATKLSFKHPVIGVHVRRTKVTGEAFHPIDEEVMVHEHFOL  
LARMQDKKHLVLTADPSLKEAKTKIPNTEFISNSISWAGLHNRVYENSRLGV  
ILDTHLSQALFLVCTFSQVCRVAYEIMQTLHPDASANFSLDDIYVFGONAHQI  
AIYAHQPTALDEIPMEPGDITGVAGNHWGYSKGVNRKLGRTGLYPSYKVRKEIETVK  
YVTPPEAK"

variation	2218	/gene="FUT8" /note="II" /frequency=".5" /replace="c" 2158..2763 /gene="FUT8" /note="putative" 2770..2775 /gene="FUT8" /note="putative" 2869..2874 /gene="FUT8" /note="putative"	905 a	597 c	663 g	827 t	
polyA_signal							
polyA_signal							
polyA_signal							
BASE COUNT	905 a	597 c	663 g	827 t			
ORIGIN							
Query Match	78.4%;	Score 1575.2;	DB 9;	Length 2992;			
Best Local Similarity	87.5%;	Pred. No. 0;					
Matches 1781;	Conservative 0;	Mismatches 223;	Indels 32;	Gaps 4;			
Qy	1	AACAGAACTTATTTTCCTGCTGGCTTAACCTAGAACACAGAGTACAAATGTTTCCAAATCTT	60				
Db	328	AACAGAAGCTTATTCACCTGTGCATTAACCTAGAAAACAGAGTTACAATGTTTTCAAATCTT	387				
Qy	61	TGAGCTCCGAGAAAGACA--GAAGGAGGATTGAAACTCTGAAAATCGGGGCATGGAAGTCTGTT	118				
Db	388	TGAGCTCCAGGACTCCAGGGAAGTGAGTTGAAAATCTGAAAATCGGGCCATGGAAGTCTGTT	447				
Qy	119	CTGGCGTTGGATTATGCTCATCTTTTTCCTGGGGACCTTATGTTTATATAGGTG	178				
Db	448	CTGGCGTTGGATTATGCTCATCTTTTTCCTGGGGACCTTGTGTTTATATAGGTG	507				
Qy	179	GTCAATTTGGTTTCGAGATATGACCCCTGACCATCTTAGCAGAGAACTCTCCAGAAATTC	238				
Db	508	GTCAATTTGGTTTCGAGATATGACCCCTGACCATCTTAGCAGAGAACTCTCCAGAAATTC	567				
Qy	239	TTGCAAAAGCTGGAGCGCTTTAAACCAACAAATTAAGAGACTTCGAGGAGAAATGGCTGAGTCTC	298				
Db	568	TGCAAAAGCTTGAACCTTTAAACCAACAAATTAAGAGACTTCGAGGAGAAATGGCTGAGTCTC	627				
Qy	299	TCCGAATACAGAAAGCCCTTATGATCAGGGGACAGCTACAGGAAGAGTCGGTGTGTTAG	358				
Db	628	TCCGGATACAGAAAGCCCTTATGATCAGGGGACAGCTATAGGAAGAGTACGGCTGTTTATG	687				
Qy	359	AAGAAAGCTTGTAAAGCCAAAGAACAGATGGAATTTACAGAAACAGAGCTAGGAATG	418				
Db	688	AAGAGAGCTTGTAAAGCCAAAGAACAGATGGAATTTACAGAAACAGAGCTAGGAATG	747				
Qy	419	ATCTGGGAAAGGATCATGAAATCTTTAAGGAGGAGGATTTGAAATTTGAGAGCTCTAAGAGCTCT	478				
Db	748	GTCTGGGAAAGGATCATGAAATCTTTAAGGAGGAGGATTTGAAATTTGAGAGCTCTAAGAGCTCT	807				
Qy	479	GGTTTTTCTTACAAAGTGAATTAAGAAATTAAGAAATTAAGAAAGGAAACGAACTCCAAA	538				
Db	808	GGTTTTTCTTACAGAGTGAATTAAGAAATTAAGAAATTAAGAAAGGAAATGAATCTCCAAA	867				
Qy	539	GACATCGAGATGAAATCTTTTGGATTTAGGACATCATGAAAGGTCTATCATGACAGATC	598				
Db	868	GACATCGAGATGAAATCTTTTGGATTTAGGACATCATGAAAGGTCTATATATGACGGATC	927				
Qy	599	TATACCTACCTCAGTCAACAGATGAGCAGAGTGTAGTGGCGGAAAGAAAGAAAGATC	658				
Db	928	TATACCTACCTCAGTCAACAGATGAGCAGAGTGTATTTGGCGGAAAGAAAGAAAGATC	987				
Qy	659	TGACAGAGCTGGTCCAGCGGAGAAATAACATATCTGAGAAATCCCAAGGACTGCGAGCAAG	718				
Db	988	TGACAGAACTGGTTTCCAGCGGAGAAATAACATATCTTCAAAATCCCAAGGACTGCGAGCAAG	1047				
Qy	719	CCAGAAAGCTGGTATGTAATATCAACAAAGGCTGTGGCTATGATGTGCACTCCCATG	778				
Db	1048	CCAGAAAGCTGGTGTATATCAACAAAGGCTGTGGCTATGATGTGCACTCCCATG	1107				

Db 2188 CAGTTCGACCAACTCAGTCAAACTTTGAGCCAACTCTAGATGAAGAGGGCTCTGA 2247  
Qy 1894 ----ACTACATATGTTTCATTCAGACATGCTCCGACCAAGAGCAAGTGGGACCCCT 1949  
Db 2248 TCTAACAAAATAAGGTATATAGTAGATATCTCTCAGCACCAAGAGCAGCTGGGAACTGA 2307  
Qy 1950 CAGATGCTGC-ACCTGGTGAACGCTCTTTTGTGAAGGGCTGCTGCGCCCTCAAGCC 2004  
Db 2308 CATAGGCTTCAATTGGTGAATCTCTTTTAAAGGGCTGCAATGCTCATACCC 2363

RESULT 9  
LOCUS HSY17978 3186 bp mRNA linear PRI 06-NOV-1998  
DEFINITION Homo sapiens mRNA for glycoprotein 6-alpha-L-fucosyltransferase  
transcript Al.  
ACCESSION Y17978  
VERSION Y17978.1 GI:3451266  
KEYWORDS FUT8 gene; glycoprotein 6-alpha-L-fucosyltransferase.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Cailleau,A., Balanzino,L., Candellier,J.J., Oriol,R. and  
Mollicone,R.  
TITLE Differential splice variants of human FUT8 embryonic cDNA  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3186)  
AUTHORS Cailleau,A.L.V.  
TITLE Direct Submission  
JOURNAL Submitted (19-AUG-1998) A.L.V. Cailleau, INSERM U178, 16 av P.V.  
Couturier, 94807 Villejuif Cedex, FRANCE  
COMMENT Related entry: D89289.  
FEATURES  
source  
1..3186  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/map="14q23"  
/clone="FUT8 transcript Al"  
/tissue\_type="whole embryo"  
/dev\_stage="40-70 days"  
1..3186  
/gene="FUT8"  
1..389  
/gene="FUT8"  
/note="A"  
390..488  
/gene="FUT8"  
/note="C"  
489..>2444  
/gene="FUT8"  
/note="D"  
717..2444  
/gene="FUT8"  
/EC\_number="2.4.1.68"  
/note="allele 1"  
/codon\_start=1  
/product="glycoprotein 6-alpha-L-fucosyltransferase"  
/protein\_id="CAA76987.1"  
/db\_xref="GI:3451267"  
/db\_xref="SPTREMBL:O00235"  
/translation="MRPWTGSRWIMLILFAWGTLLFYIGHLVRDNDHPDHSRELS  
KILAKLEIKQNEDLRMAESLRIPGPIDQPAIGRVRLVLEQLVKAQEIENYKK  
QTRNGLKDHETILRPERENGAKLWFLQSELKKLKNLEGNELORHAEFLLDI-GHHE  
RSIMTDLYLQSDGAGDWRKEAKDLTELQRRITYLQNPDKCSKAKLVCNINKGC  
GYGCOLHHVYCFMAYGTQRTILLESQNRWYATGWETVFRPVSCTCTDSRGISTGH  
WSGEYKQNVQVLPIDVSLHPPPLPLAVPEDLADLRVHGDPAVMWVSQFVKY  
LIRPOPWEKIEEATKKLGFPHVIGVHVRRTDKVTEAAFPHEEYVMVHVBHFOL  
LARRMQVDKRVYLATDPSLLKEAKTKYPNVEFISDINSISWSAGLHNRNTENSILRGV

ILDIHLFSQADFLVCTESSQVRVAYEIMQTLHPDASANFHSLLDIYFPGQNAHQI  
AIYAHOPUTADEIPMEPDITGVAGNHWGYSKGNVRKLGRTGLYPSYKVRKIEITVK  
YPTVPEAKK  
2506  
/gene="FUT8"  
/note="11"  
/frequency=".5"  
/replace="c"  
3046..3051  
/gene="FUT8"  
/note="putative"  
3058..3063  
/gene="FUT8"  
/note="putative"  
3157..3162  
/gene="FUT8"  
/note="putative"  
BASE COUNT 910 a 710 c 721 g 840 t 5 others  
ORIGIN  
Query Match 78.4%; Score 1575.2; DB 9; Length 3186;  
Best Local Similarity 87.5%; Pred. No. 0;  
Matches 1781; Conservative 0; Mismatches 223; Indels 32; Gaps 4;  
Qy 1 AACGAAACTTATTTCTGTGTGCTAACTAGAACAGAGTACAATGTTTCCAACTCTT 60  
Db 616 AACGAAAGTCTATTCACTGTGCACCTAACTAGAACAGAGTTACAATGTTTCAATCTT 675  
Qy 61 TGAGCTCGGAGAGACA-GAAGGAGTTGAACTCTGAAATCGGGCATGGACTGTT 118  
Db 676 TGAGCTCCAGACTCCAGGGAAGTGAGTTGAAATCTGAAATCGGGCATGGACTGTT 735  
Qy 119 CTTGCGGTGATTAATGCTCAATCTTTTCTGCTGGGACCTTATTTGTTTATATAGTG 178  
Db 736 CTTGCGGTGATTAATGCTCAATCTTTTCTGCTGGGACCTTCTGCTGTTTATATAGTG 795  
Qy 179 GTCAATTTGGTTCGAGATAATGACCACTGACCATCTCTAGCAGAGAACTCTCCAGATTC 238  
Db 796 GTCACTTGGTACGAGATAATGACCATCTCTGATCACTCTAGCCGAGAACTGTCCAGATTC 855  
Qy 239 TTGCAAGCTGGAGCGCTTAAACAAACAAATGAAGACTTTGAGGAGAAATGCTGAGTCTC 298  
Db 856 TGGCAAGCTTTGAACGTTTAAACACAGCAGATGAAGACTTTGAGGCGAATGGCCGAATCTC 915  
Qy 299 TCCGAATACCAAGAGGCTCTATTGATCAGGGGACAGCTTACAGGAGAGTCCGCTGTTTAG 358  
Db 916 TCCGGATACCAAGAGGCTCTATTGATCAGGGGACAGCTTATAGGAGAGTACCGCTTTTAG 975  
Qy 359 AAGAACAGCTTTGTTAAAGCCCAAGAACAGATTGAAATTTACAGAAACCAAGCTAGGAATG 418  
Db 976 AAGAGCAGCTTGTAAAGCCCAAGAACAGATTGAAATTTACAGAAACCAAGCTAGGAATG 1035  
Qy 419 ATCTGGAAAGGATCATGAATCTTTAAGGAGAGAGATTGAAATTTGAGAGCTTAAGAGCTCT 478  
Db 1036 GTCTGGGAGAGGATCATGAATCTCTGAGGAGAGAGATTGAAATTTGAGAGCTTAAGAGCTCT 1095  
Qy 479 GGTCTTTCTACAAAGTGAATTTGAAGAAATTAAGAAATTAAGAAACCAAGCTCCAAA 538  
Db 1096 GGTCTTTCTACAGAGTGAATTTGAAGAAATTAAGAAATTAAGAAACCAAGCTCCAAA 1155  
Qy 539 GACATGACAGTGAATTTCTTTTGGATTAGGACATCATGAAAGGTCTATCATGACAGATC 598  
Db 1156 GACATGACAGTGAATTTCTTTTGGATTAGGACATCATGAAAGGTCTATATGACGGATC 1215  
Qy 599 TATACCTACCTCAGTCAACACAGATGGAGCAGGTGAGTGGCGGAAACCAAGCTAGGAATG 658  
Db 1216 TATACCTACCTCAGTCAACACAGATGGAGCAGGTGAGTGGCGGAAACCAAGCTAGGAATG 1275  
Qy 659 TGACAGAGCTGGTCCAGUGGAGATAACATATCTGCAGAAATCCCAAGACTCCGACCAAG 718  
Db 1276 TGACAGAACTGGTTTTCAGCGGAGATAACATATCTTCAGAAATCCCAAGACTCCGACCAAG 1335  
Qy 719 CCAGAAAGCTGGTATGTAATATCAACAAAGGCTGTGGCTATGGATGTCAACTCCATCATG 778







QY 1799 ACCCTACATATCTCTGAAGCTGAAATAAGATGGAGTGTAAAGATTAA----- 1848  
Db 2416 ACCCCACATATCTCTGAGGCTGAGAAATAAGCTCAGATGAAGAGATAAACGACCAAACT 2475  
QY 1849 -----CAACAGAAATTAGTTCAGACCATCTCAGCAAGCAGAGACCAG----- 1893  
Db 2476 CAGTTCCGACCAAACTCAGTTCAAAACCATTTGAGCCAAACTGTAGATGAAGAGGGCTCTGA 2535  
QY 1894 ----ACTACATATGTTTCATTGACAGACATCTCGCACCAAGAGCAAGTGGGAACCT 1949  
Db 2536 TCTAACAAATAAGGTTATAGTAGATATCTCAGCACCAAGAGCAGCTGGGAACCTGA 2595  
QY 1950 CAGATGCTGC-ACTGGTGAAGCGCTCTTTGTGAAGGGCTGTGTGCCCTCAAGCC 2004  
Db 2596 CATAGGCTTCAATTGGTGAATTCCTCTTTTACAGAGGGCTGCAATGCCTCATACCC 2651

RESULT 11  
HSA536055  
LOCUS  
DEFINITION Homo sapiens mRNA for alpha6-fucosyltransferase (FUT8 gene), splice variant B5.  
ACCESSION AJ536055  
VERSION AJ536055.1 GI:27552399  
KEYWORDS alpha6-fucosyltransferase; FUT8 gene.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
Mollicone, R., Michalewski, J.C., Bauvy, C., Cailletau-Thomas, A., Candelier, J.J., Martinez-Duncker, I., Breton, C., Codogno, P. and Oriol, R.  
TITLE Splice variants of alpha6-fucosyltransferase are expressed early in human embryogenesis  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3568)  
AUTHORS Oriol, R.  
TITLE Direct Submission  
JOURNAL Submitted (06-JAN-2003) Oriol R., U504, INSERM, 16 Avenue Paul Vaillant-Couturier, Villejuif, 94807, FRANCE  
COMMENT related splice variants AJ536053.1, AJ536054.1 and AJ536056.1.  
FEATURES  
Location/Qualifiers  
1..3568  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/map="14q23.2"  
/clone="B2D-L1"  
/tissue\_type="whole embryo"  
/dev\_stage="40-70 days"  
/country="France:Paris"  
1..3568  
/gene="FUT8"  
1098..2825  
/gene="FUT8"  
/EC\_numbers="2.4.1.68"  
/function="adds fuc in alpha1,6 to the internal GlcNAc of the core chitobiose"  
/codon\_start=1  
/evidence=experimental  
/product="alpha6-fucosyltransferase"  
/protein\_id="CAD59926.1"  
/db\_xref="GI:27552400"  
/translation="MRPVTGSRWIMLILFAWGTLLFYIGHVLVRDNDHPDHSRELS KILAKLERKQNEDLRMAESLRIPGPIDQGAIGRVRVLEQLVKAKEQIENYKK QTRNLGKHDEILRRRIENGAKELWFLQSELKKLLEGNELORHAFELLDIGHHE RSIWTDLYLSQTAGDWREKAKDLTELVRRTIYLNQPKDCSKAKLVCNINKGC GYGCOLHNVCFMAYGTQRTILLESQNRWRYATGWTVPFRPSETCTDRSGISTGH WSGEVKDNVQVLEPFI VDSLHPRPPLPLAVPDLADLRVVRVHGDPAVWVMSQFYKY LIRPQWLEKSEI EATKKLGFKHPVIGVHVVRRTDKVGTAAAFHPFIEYVMVHEHFOL

gene  
CDS

LARMQVQKRVYLATDPDLLKEAKTKYPNYEFISDINSWSAGLHNRNTENSIRGV  
ILDHFLSQADFLVCTFSQVRVAYEIMQTLHPDASANFHSLLDDIYFFGQNAHQI  
AIYAHQPRTADEIPMEPGDIIGVAGNHWGYSKGVNRKLGRTGLYPSYKVRKREITVK  
VPTYPEAEK"  
3538..3543  
/gene="FUT8"  
3553  
/gene="FUT8"  
BASE COUNT 1054 a 751 c 788 g 975 t  
ORIGIN  
Query Match 78.4%; Score 1575.2; DB 9; Length 3568;  
Best Local Similarity 87.5%; Pred. No. 0;  
Matches 1781; Conservative 0; Mismatches 223; Indels 32; Gaps 4;  
QY 1 AACAGAAACTTATTTTCTGTGTGGCTAACTAGAACACAGAGTACAAATGTTTCCAAATCTT 60  
Db 997 AACAGAACTTATTCACCTGTGCACTAACTAGAAACAGAGTTACAAATGTTTCAATCTT 1056  
QY 61 TGAGCTCCGAGAGACACA--GAAGGGAGTTGAAACTCTGAAATCGGGCATGGACTGGTT 118  
Db 1057 TGAGCTCCAGGACTCCAGGGAAGTGAGTTGAAATCTGAAATCGGGCATGGACTGGTT 1116  
QY 119 CTTGGCGTTGAGTATGCTCATCTTTTTCCTGGGGACCTTATTTGTTTATATAGTG 178  
Db 1117 CTTGGCGTTGAGTATGCTCATCTTTTTCCTGGGGACCTTATTTGTTTATATAGTG 1176  
QY 179 GTCAATTTGTTTCGAGATAATGACCACCTGACCACTTCTAGCAGAGAACTCTCCAAAGATTC 238  
Db 1177 GTCAATTTGTTTCGAGATAATGACCACCTGACCACTTCTAGCAGAGAACTCTCCAAAGATTC 1236  
QY 239 TTGCAAAAGCTGGAGCGCTTAAACCAACAAATTAAGAATTTGAGGAGAAATGGCTGAGTCTC 298  
Db 1237 TGCAAAAGCTTGAACGCTTAAACCAAGAGTGAAGACTTGAGGCGAATGGCCGAATCTC 1296  
QY 299 TCCGAATACCAAGAGGCGCTTATGATCAGGGGACAGCTCAGAGAGAGTCCGTTGTTTAG 358  
Db 1297 TCCGATACCAAGAGGCGCTTATGATCAGGGGACAGCTTATGAGGAGAGTACGCGCTTTAG 1356  
QY 359 AAGAACAGCTGTTTAAAGCCAAAGACAGATTTGAAATTTACAAAGAACTAGGAGTATG 418  
Db 1357 AAGAACAGCTGTTTAAAGCCAAAGACAGATTTGAAATTTACAAAGAACTAGGAGTATG 1416  
QY 419 ATCTGGAAAGGATCATGAAATCTTAAAGGAGGAGTGAATAATGGAAGTGAAGAGCTCT 478  
Db 1417 GTCTGGGGAAGGATCATGAAATCTTGAAGGAGGAGTGAATAATGGAAGTGAAGAGCTCT 1476  
QY 479 GGTCTTTTCTACAAAGTGAATTTGAAGAAATTAAGAAATTTAGAGGAAACGAAGTCCAA 538  
Db 1477 GGTCTTTTCTACAAAGTGAATTTGAAGAAATTTAGAGGAAATTAAGAACTGAAGTCCAA 1536  
QY 539 GACATGAGATGAAATTTCTTTTGGATTTAGACATCATGAAAGGTCTATCATGACAGATC 598  
Db 1537 GACATGAGATGAAATTTCTTTTGGATTTAGACATCATGAAAGGTCTATATATGACGATC 1596  
QY 599 TATATACCTCAGTCAAAACAGATGGAGCAGGTGAGTGGCGGGAAGAAAGAGCCAAAGATC 658  
Db 1597 TATATACCTCAGTCAAAACAGATGGAGCAGGTGAGTGGCGGGAAGAAAGAGCCAAAGATC 1656  
QY 659 TGACAGAGCTGGTCCAGCGAGGAATAACATATCTGAGAGATTTCCAGAGCTGCGAGCAAG 718  
Db 1657 TGACAGAACTGGTTTTCAGCGAGGAATAACATATCTTCAAGAAATCCCAAGGACTGCGAGCAAG 1716  
QY 719 CCAGAAAGCTGGTATGTAATTAATCAACAAAGGCTGTGGCTATGAGTGTCAACTCCATCATG 778  
Db 1717 CCAGAAAGCTGGTGTGTAATTAATCAACAAAGGCTGTGGCTATGAGTGTGAGTGTGATG 1776  
QY 779 TGGTTTACTGCTTCATGATTTGCTTATGGCAGCCAGCAACATCATCTTGGAAATCTCAGA 838  
Db 1777 TGGTTTACTGCTTCATGATTTGCATATGGCAGCCAGCAACATCATCTTGGAAATCTCAGA 1836  
QY 839 ATTGGCGCTATGCTACTGAGGAGTGGAGAGCTGTTTATAGACCTGTAAAGTGAAGCATGCA 898



1837	ATTGGGCGCTATGCTATCTGGTGGATGGGAGACTGTATTTAGGCCCTGTAAAGTGAGACATGCA	1899
Qy	CAGACAGGCTCTGGCCTCTCCACCTGGACACTGGTCAAGTGAAGGACAAAAATGTTTC	958
Db		
1897	CAGACAGATCTGGCATCTCCACCTGGACACTGGTCAGGTGAAGTGAAGGACAAAAATGTTTC	1956
Qy	959 AAGTGGTCGAGCTCCCATTTGTFAGACAGCCTCCATCTCGTCTCTCTTACTTTACCCCTTGG	1018
Db		
1957	AAGTGGTCGAGCTTCCCATTTGTFAGACAGCTCTTCATCCCGTCTCTCCATATTTTACCCCTTGG	2016
Qy	1019 CTGTACCAGAAAGACCTTTGCAGATCGACTCTGTAGAGTGCATGTGTGATCTCTGCAGTGTGGT	1078
Db		
2017	CTGTACCAGAAAGACCTCGCAGATCGACTTGTACGAGTGCATGTGTGACCTCTGCAGTGTGGT	2076
Qy	1079 GGGTATCCCAAGTTTGTCAAAATACTTGATCCGCTCCACAACCTTGGCTGGAAAGGAAATAG	1138
Db		
2077	GGGTGTCTCAGTTTGTCAAAATACTTGATCCGCTCCACAACCTTGGCTAGAAAAAGAAATAG	2136
Qy	1139 AAGAAACCAACAAGAGCTTGGCTTTCAAACATCGAGTTATTGGAGTCCATGTCCAGAGCGCA	1198
Db		
2137	AAGAAAGCCACAAGAGCTTGGCTTCAAAACATCCAGTTATTGGAGTCCATGTCCAGAGCGCA	2196
Qy	1199 CTGCAAAAGTGGGAACAAGAGCAGCTTCCATCCCATTTGAGGAGTAATCATGTGTACACGTTTG	1258
Db		
2197	CAGCAAAAGTGGGAACAAGAGCTGCTTCCATCCCATTTGAAGAGTGCATGGTGCAATGTTG	2256
Qy	1259 AAGAAACATTTTACGCTTCTCGAAGCGCAGAAATGAAGTGGATAAAAAAGAGTGTATCTCG	1318
Db		
2257	AAGAACATTTTACGCTTCTTGGACGCGCAGAAATGCAAGTGGACAAAAAGAGTGTATTTGG	2316
Qy	1319 CCACGTGATGACCCCTTTCTTTGTTTAAAGAGGCAAGACAAGAGTACTCCAAATATATGAATTTA	1378
Db		
2317	CCACAGATGACCCCTTCTTTATTAAAGAGGCAAAAAACAAGTAGTACCCCAATATATGAATTTA	2376
Qy	1379 TTAGTGATAACTCTATTTCTTTGGTCAGCTGGACTACACAAACCGATACACAGAAAAATTCAC	1438
Db		
2377	TTAGTGATAACTCTATTTCTTGGTCAGCTGGACTGCAAAATCGATACACAGAAAAATTCAC	2436
Qy	1439 TTGGGGCGGTGATCCCTGGATATACACTTTTCTCTCCAGGCTGACTTCTCTTGTGTGTACTT	1498
Db		
2437	TTCTGTGGAGTGATCTTGGATATACATTTTCTCTCAGGCGAGACTTCTTAGTGTGTACTT	2496
Qy	1499 TTTTCATCCCAAGGCTGTAGGGTTGCTTTATGAANAATCATGCAAAACACTGCATCTCTGATGCCT	1558
Db		
2497	TTTTCATCCCAAGGCTGTGCGAGTTCCTTTATGAAATTTATGCAAAACACTACATCTCTGATGCCT	2556
Qy	1559 CTGCAAACTTCCATCTTTAGATGACATCTACTATTTTGGAGGCCAAAAATGCCCAACACC	1618
Db		
2557	CTGCAAACTTCCATCTTTTATAGATGACATCTACTATTTTGGGGGCCAGAAATGCCCAACAATC	2616
Qy	1619 AGATTGACAGTTTATCCTCCACCAACCTCGAACTAAAGAGGAAATCCCCATCGGAACCTGGAG	1678
Db		
2617	AAATTGCCATTTATGCTTCAACCAACCCGAACTGCAGATGAATTTCCCATCGGAACCTGGAG	2676
Qy	1679 ATATCATTTGGTGTGGCTGGAAACCAATTTGGAATGGTTACTCTTAAAGGTGTCAACAGAAAAAC	1738
Db		
2677	ATATCATTTGGTGTGGCTGGAAATCATTTGGGATGGCTATTTCTTAAAGGTGTCAACAGGAAAT	2736
Qy	1739 TAGAAAAACAGGCCTGTGACCTTCTCTACAAAGTCCGAGAGAGATAGAAACAGTCAAT	1798
Db		
2737	TGGGAAGGACGGGCTATATCCCTCTCTACAAAGTTCGAGAGAGATAGAAACCGTCAAGT	2796
Qy	1799 ACCCTACATATCTCGAGCTGAAAAATAGAGATGGAGTGTAAAGAGATTAA-----1848	
Db		
2797	ACCCCAACATATCTCTGAGGCTCGAANAATAAAGCTCAGATGGAAGAGATAAACGACCAAACT	2856
Qy	1849 -----CAACAGAAATTTAGTTTCAAGACCATCTCAGGCCAAGCAGAACCCAG-----1893	
Db		
2857	CAGTTCGACCAAACTCAGTTTCAAAACCATTTGAGCCAAACTGTAGATGAAGGGCTCTGTA	2916
Qy	1894 -----ACTAACATATGGTTTCAATTGACAGACATGCTCCGCAACAAGAGCAAGTGGGAACCTT	1949
Db		
2917	TCTAACAAAAATAAGGTTATATGAGTAGATACTCTTCAGCAACAAGAGCAGCTGGGAATCTGA	2976

Qy 1950 CAGATGCTGC-ACTGTGGAAACGCCCTTTTGTGAAGGCGTGCTGTGCGCCTCAAGCC 2004  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 2977 CATAGGCTTCATTGTGGTAATTCCTCTTTAAACAAGGCGTCAATGCCTCATACCC 3032  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12  
HSA536053 3666 bp mRNA linear PRI 08-JAN-2003  
LOCUS Homo sapiens mRNA for alpha6-fucosyltransferase (FUT8 gene), splice  
DEFINITION variant B3.  
ACCESSION AJ536053  
VERSION AJ536053.1 GI-27552395  
KEYWORDS alpha6-fucosyltransferase; FUT8 gene.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1  
AUTHORS Mollicone,R., Michalski,J.C., Bauvy,C., Caillean-Thomas,A.,  
Candelier,J.J., Martinez-Duncker,I., Breton,C., Codogno,P. and  
Oriol,R.  
TITLE Splice variants of alpha6-fucosyltransferase are expressed early in  
human embryogenesis  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3666)  
AUTHORS Oriol,R.  
TITLE Direct Submission  
JOURNAL Submitted (06-JAN-2003) Oriol R., U504, INSERM, 16 Avenue Paul  
Vaillant-Couturier, Villejuif, 94807, FRANCE  
COMMENT related splice variants AJ536054.1, AJ536055.1 and AJ536056.1.  
FEATURES Location/Qualifiers  
source  
1..3666  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/map="14q23.2"  
/clone="B2CD-L1"  
/tissue\_type="whole embryo"  
/dev\_stage="40-70 days"  
/country="France.Paris"  
1..3666  
/gene="FUT8"  
1196..2973  
/gene="FUT8"  
/EC\_number="2.4.1.68"  
/function:"adds fuc in alpha1,6 to the internal GlcNAc of  
the core chitobiose"  
/codon\_start=1  
/evidence=experimental  
/product="alpha6-fucosyltransferase"  
/protein\_id="CAD59924.1"  
/db\_xref="GI:27552395"  
/translation="MRPWTGSWRWMLILFAWGTLTFYIGHLVRDNDHPDHSHRELSL  
KILAKRLRQNEDELRMAESLRIPGPIDQGPAIGRVLEQLVKAQEIKENYKKR  
RTNQDLKHDEIRRIENGAKELWFPELSEKLKNLEGNELQRHADFLDLGLHGHE  
RSIMTDLYLSOTDGRGDWEKBAKDITELVORRI TYLONPKDCSKAKLVCNINKNG  
GYGCOLHHVVYCFMIAVTQTLLILESONRYATGWETVPRPVSETCTDRSGISTGH  
WSGEPKNNQVVEALPIVDLSLHPPPYLPPLAVPEDLDLRLVRVHGDPAVMVVSQFVK  
LI RPQWLKEIIEATKKGKGFHPVI GVHVRRTDKVGTEAFHPIETMVHVEHFQKI  
LARMQVDKRVVLATDDPSLLEAKTKPNYEFISDNISWSAGLNHRVTENSLRGV  
ILDTHFSQADFLVCFTSSQCVRAYEIMOTLHPDASANPHSLDDIYYFGQNAAHQI  
AIVAHOPRTADEIMPEDDIIGVAGNHWDGYSKVNRKLGRTGLGPSYKVKREKTETVK  
YPYTPAEAK"

polyA\_signal 3636..3641  
/gene="FUT8"  
polyA\_site 3651  
/gene="FUT8"

BASE COUNT 1077 a 768 c 808 g 1013 t  
ORIGIN

Query Match .784%; Score 1575.2; DB 9; Length 3666;



VERSION	AJ536054.1	GI:27552397	
KEYWORDS	alpha6-fucosyltransferase; FUT8 gene.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 Mollicone, R., Michalski, J.C., Bauvy, C., Caillet, Thomas, A., Candelier, J.J., Martinez-Duncker, I., Breton, C., Codogno, P. and Oriol, R.		
TITLE	Splice variants of alpha6-fucosyltransferase are expressed early in human embryogenesis		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 3772)		
AUTHORS	Oriol, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (08-JAN-2003) Oriol R., US04, INSERM, 16 Avenue Paul Vaillant-Couturier, Villejuif, 94807, FRANCE		
COMMENT	related splice variants AJ536053.1, AJ536055.1 and AJ536056.1.		
FEATURES	Location/Qualifiers		
source	1. .3772		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/chromosome="14"		
	/map="14q23.2"		
	/clone="B2CD-L2"		
	/tissue_type="whole embryo"		
	/dev_stages="40-70 days"		
	/country="France:Paris"		
gene	1. .3772		
	/gene="FUT8"		
CDS	1196..2923		
	/gene="FUT8"		
	/EC_number="2.4.1.68"		
	/function="adds fuc in alpha1,6 to the internal GlcNAc of the core chitobiose"		
	/codon_start=1		
	/evidence=experimental		
	/product="alpha6-fucosyltransferase"		
	/protein_id="CAD59925.1"		
	/db_xref="GI:27552398"		
	/translation="MRPWTGSRWIMLILFAWGLTLFVIGHLVRDNDHPHSRELS KILAKRLKQNDLRMAESLRIPEGIDQGPAGRVRLBOLVKAKEQIENYKK OTRNGLSKHDLRRLRENGAKELWFLQSELKLNLEGNELORHAEFLDLGHE RSIMIDLFLSOTGDGWRKEAKDLTELVRITLYLONPKOCSKAKLVNINKGC GYCOLHHVYCFMIATGYQRTLILESQNNRYATGMEVFPVSETCTDRSGISTG WSGVKQKNVQVLEPLIVDSLPRPPYPLAVPEDLADRLVRHGDPAVMWSQFVKY LIRPQMLEKEIEBATKLGPKHPVIGVHVNRITDKVTEAAPHPIESYHVHVEHFL LARMOVDKRVYLATDDPSLLKEAKTYPNVEFISNLSWSAGLHNRVTENSLRGV ILDIHFLSQADFLVCTPSSQVCRVAYEIMQTLHPDASANFHSLLDDIYVFGONAHNOI AIYAHQPTADEIPEPEGDIIGVAGNHWGYSKGNVRKLGRTGLYPSYKVKREKLETVK YPTYPEAEK"		
polyA_signal	3723..3728		
	/gene="FUT8"		
polyA_site	3757		
	/gene="FUT8"		
BASE COUNT	1110 a 777 c 830 g 1055 t		
ORIGIN			
Query Match	78.4%;	Score 1575.2;	DB 9; Length 3772;
Best Local Similarity	87.5%;	Pred. No. 0;	
Matches 1781;	Conservative 0;	Mismatches 223;	Indels 32; Gaps 4;
Qy	1 AACAGAACTTATTTCTGTGTGGCTAACTAGAACACAGAGTACAAATGTTTCCAAATCTT 60		
Db	1095 AACAGAACTTATTTCACTGTGCACTAAGTAAACAGAGTACAAATGTTTCCAAATCTT 1154		
Qy	61 TGAGCTCCGAGAGACA--GAAGGAGTTGAACTCTGAAAATCGCGGCATGGACTGGTT 118		
Db	1155 TGAGCTCCGAGACTCCAGGAGAGTGAGTTGAAAATCTGAAAATCGCGGCATGGACTGGTT 1214		
Qy	119 CCTGCGGTGGATTATGCTATCTTTTGGCTGGGGACCTTATTTGTTTATATAGTGT 178		

Db 2295 CAGACAAAGTGGGAACAGAAAGCTGCCCTCCATCCCATTTGAAGAGTACATGGTGCATGTTG 2354  
Qy 1259 AAGAACAATTTTCAGCTTCTCGAACGAGAAATGAAGTGAATAAAAAAGAGTGTATCTGG 1318  
Db 2355 AAGAACAATTTTCAGCTTCTCGAACGAGAAATGAAGTGAATAAAAAAGAGTGTATTTGG 2414  
Qy 1319 CACATGATGACCTTCTTGTGTTAAAGGAGGAAAGACAAAGTACTCCAAATATGAAATTA 1378  
Db 2415 CCACAGATGACCTTCTTGTGTTAAAGGAGGAAAGACAAAGTACTCCAAATATGAAATTA 2474  
Qy 1379 TTAGTGATAACTCTATTTCTCGTGCAGCTGGACTACAAACGATACACAGAAATTCAC 1438  
Db 2475 TTAGTGATAACTCTATTTCTCGTGCAGCTGGACTACAAACGATACACAGAAATTCAC 2534  
Qy 1439 TTGGGGGCGTATCTCGATATACACTTTCTCTCCAGGCTGACATCTCTTGTGTACTTT 1498  
Db 2535 TTGGTGGAGTATCTCGATATACACTTTCTCTCCAGGAGTCTCTAGTGTACTTT 2594  
Qy 1499 TTTCATCCAGCTCTGTAGGTTGCTTATGAATCATGCAACACTGCACTCTGATGCT 1558  
Db 2595 TTTCATCCAGCTCTGTAGGTTGCTTATGAATCATGCAACACTGCACTCTGATGCT 2654  
Qy 1559 CTGCAAACTTCCATTTTAGATGACATCTACTATTTGGAGGCCAAATGCCACAACC 1618  
Db 2655 CTGCAAACTTCCATTTTAGATGACATCTACTATTTGGAGGCCAAATGCCACAATC 2714  
Qy 1619 AGATTGCACTTTATCTCACCACCTCGAACTTAAAGAGGAAATCCCATGGAACCTGGAG 1678  
Db 2715 AATTGGCCATTTATGCTCACCACCCGACCTGAGATGAAATTTCCCATGGACCTGGAG 2774  
Qy 1679 ATATCATTTGGTGTGGTGAACCAATTTGAATGTTACTCTTAAAGTGTCAACAGAAAC 1738  
Db 2775 ATATCATTTGGTGTGGTGAACCAATTTGAATGTTACTCTTAAAGTGTCAACAGAAAT 2834  
Qy 1739 TAGGAAACAGCGCTGTACCTCTCAAGTCCGAGAGAGATAGAAACAGTCAAT 1798  
Db 2835 TGGGAAGGAGCGGCTTATATCCCTCTCAAGTTCGAGAGAGATAGAAACGGTCAAGT 2894  
Qy 1799 ACCCTACATATCTGAAGCTGAAATAGAGATGAGTGTGAAGATTTAA----- 1848  
Db 2895 ACCCACATATCTGAGCTGAGAAATTAAGCTCAGATGGAAGATTAACGACCAACT 2954  
Qy 1849 -----CAACAGATTTAGTTCAGACATCTAGCCAGCAGAGACCCAG----- 1893  
Db 2955 CAGTTCCGACCAACTCAGTTCAACCAATTTGAGCCAAACTGTAGATGAAGGGCTCTGA 3014  
Qy 1894 ----ACTACATATGTTTCATTTGACAGATGCTCCGACCAAGAGCAAGTGGGACCT 1949  
Db 3015 TCTAACAAATAAGTTATATGATAGATATCTCTCAGCACCAGAGCAGCTGGGAATCGA 3074  
Qy 1950 CAGATGCTGC-ACTGGTGAAGCGCTCTTTGTGAAGGCTGTGTGCCCTCAAGCC 2004  
Db 3075 CATAGGCTCAATGGTGGATTCTCTTTAAAGGGCTGCAATGCTCATACCC 3130

RESULT 14  
HSA536056  
LOCUS HSA536056 4196 bp mRNA linear PRI 08-JAN-2003  
DEFINITION Homo sapiens mRNA for alpha6-fucosyltransferase (FUT8 gene), splice variant B6.  
ACCESSION AJ536056  
VERSION AJ536056.1 GI:27552401  
KEYWORDS alpha6-fucosyltransferase; FUT8 gene.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Mollicone, R., Michalek, J.C., Bauvy, C., Caillaud-Thomas, A., Candelier, J.J., Martinez-Duncker, I., Breton, C., Codogno, P. and Oriol, R.  
TITLE Splice variants of alpha6-fucosyltransferase are expressed early in human embryogenesis

Unpublished  
2 (bases 1 to 4196)  
REFERENCE Oriol, R.  
AUTHORS Direct Submission  
TITLE Submitted (06-JAN-2003) Oriol, R., U504, INSERM, 16 Avenue Paul  
JOURNAL Vaillant-Couturier, Villejuif, 94807, FRANCE  
COMMENT related splice variants AJ536053.1, AJ536054.1 and AJ536055.1.  
FEATURES Location/Qualifiers  
1..4196  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/map="14q32.2"  
/clone="B3CD-L1"  
/tissue\_type="whole embryo"  
/dev\_stage="40-70 days"  
/country="France:Paris"  
1..4196  
/gene="FUT8"  
1726..3453  
/gene="FUT8"  
/EC\_number="2.4.1.68"  
/function="adds fuc in alpha1,6 to the internal GlcNAc of the core chitobiose"  
/evidence="experimental"  
/product="alpha6-fucosyltransferase"  
/protein\_id="CAD59927.1"  
/db\_xref="GI:27552402"  
/translation="MRPWTGSWIMLIFAWGTLFLFYIGHLVRDNDHPDHSRELS  
KILAKLERQKQEDLRMAESLRIPGPDQPAIGRVLRVEQLVKAKEIENYKK  
QTNRGLKDEHILRRRIENGAKELWPFQSEKLKLEGNLEQLRDEFLDLGHE  
RSIMTDLYLSQTDGAGDREKADILTEVQRRITVLPNPKDCSKAKLVNINKGC  
GSCGLHHVYCFMIAVGTORTLILSONRWYATGWTVPVPSVETCTDPSGISTGH  
WSEVKDKNVQVELPIVDSLHPRPVPLAIPEDLADRLVRVHGDPAVWVVSQFVKY  
LIRPQWLEKEIBEATKLGKHPVIGVHVRDVKVGTAAAFPIEPIYVHVEHFQL  
LAARMQVKKRVYVLTADTDPSSLKEAKTYPIYEFISDINSWSAGLHNRVTENSLRG  
ILDIHFLSQADEFVCTFSQCRVAYEIMQTLHPDASANFHSLDIYIFGCGNAHQI  
AIYAHOPRTADEIPMEPGDIIGVAGNHWDGYSKGVNRKLGRTGLYPSYKVKETETVK  
YPTVPEAEK"  
4166..4171  
/gene="FUT8"  
polyA\_signal  
4181  
polyA\_site  
/gene="FUT8"  
BASE COUNT 1245 a 854 c 900 g 1197 t  
ORIGIN  
Query Match 78.4%; Score 1575.2; DB 9; Length 4196;  
Best Local Similarity 87.5%; Pred. No. 0;  
Matches 1781; Conservative 0; Mismatches 223; Indels 32; Gaps 4;  
Qy 1 AACAGAACTATTTTCTGTGGCTAACTAGAACACAGAGTACAATGTTTCCAAATCTT 60  
Db 1625 AACAGAACTATTTCACTGTGCACCTAAGAACAGAGTTACAAATGTTTCAATCTT 1684  
Qy 61 TGAGCTCCGAGAGACACA--GAAGGAGGTGAACACTCTGAAAATCGCGGCATGGACTGTT 118  
Db 1685 TGAGCTCCAGACTCCAGGGAAGTGAAGTGAATAATCTGAAAATCGCGCCATGGACTGTT 1744  
Qy 119 CTGGCGCTGGATTATGCTCATTTCTTTTCTGGGGGACCTTATTTGTTTATATAGTG 178  
Db 1745 CTGGCGCTGGATTATGCTCATTTCTTTTCTGGGGGACCTTCTGCTGTTTATATAGTG 1804  
Qy 179 GTCAATTTGTTTCAGATAATGACCCCTGACCATTTACGACAGAACTCTCCAAGATTC 238  
Db 1805 GTCACTGGTACAGATAATGACCATCTCTATCTACTAGCCGAGAACTGTCACAGATTC 1864  
Qy 239 TTGCAAGCTGGAGCGCTTAAAAACAAACAAATGAAGACTTTGAGAGAAATCGTGAGTCTC 298  
Db 1865 TGGCAAGCTTGAACGCTTAAAAACAGCAGAAATGAAGACTTTGAGCGAAATGGCCGAATCTC 1924  
Qy 299 TCGAATACAGAAAGGCCCTATTGATCAGGGGACAGCTACAGGAGAGTCCGTGTTTAG 358

[illegible]

Db	3005	TTAGTGATAACTCVAATTTCTCGTGCAGCTGGACCTGCACAATCGATACACAGAAAATTCAC	306
Qy	1439	TTGGGGCGTGATCTCGGATATACACTTTCTCTCCAGGCTGACTTCTTGTGTGTACTT	1498
Db	3065	TTCTGTCGAGTGATCTGGATATACATTTTCTCTCTCAGGCAGACTTCTTCTAGTGTGTACTT	3124
Qy	1499	TTTTCATCCGAGTCTGTAGGGTGTCTTATGAATCATGCAAACTGCATCTCGATCTCGATGCGCT	1558
Db	3125	TTTTCATCCGAGTCTGTGAGTGTCTTATGAATTTATGCAAACTGCATCTCGATGCGCT	3184
Qy	1559	CTGCARAACTTCCATTTCTTTAGATGACATCTACTATTTTGGAGGCCAAAATGCCCAACAAC	1618
Db	3185	CTGCARAACTTCCATTTCTTTAGATGACATCTACTATTTTGGGGGCCAGATGCCCAACAAC	3244
Qy	1619	AGATTGCGAGTTTATCTCTCACCAACTCGAACTTAAAGAGGAAAATCCCATGGAACCTGGAG	1678
Db	3245	AAATTGCCATTTATGCTCTCACCAACCCCGAACTGCAGATGAAATTTCCCATGGAACCTGGAG	3304
Qy	1679	ATATCATTTGGTGTGCTGGAAACCATTTGGAATGGTTACTCTTAAAGGTGTCAACAGAAAAC	1738
Db	3305	ATATCATTTGGTGTGCTGGAAACCATTTGGAATGGTTACTCTTAAAGGTGTCAACAGGAAAT	3364
Qy	1739	TAGGAAAAACAGGGCTGTGACCTCTCTACAAAGTCCGAGAGAGATAGAAAACAGTCAAAAT	1798
Db	3365	TGGAGAGACGGGCTATATCTCTCTCAAAAGTTCGAGAGAGATAGAAAACAGTCAAGT	3424
Qy	1799	ACCTCATATCTCTGAAGCTGAAAATATAGATGGAGTGTAAAGAGATTAA-----	1848
Db	3425	ACCCACATATCTCTGAGGCTGAGAAATAAGCTCAGATGGAAGAGATAAACGACCAAACT	3484
Qy	1849	-----CAACAGAAATTTAGTTCAGACCAATCTCAGCCAGCAGCAGACCCAG-----	1893
Db	3485	CAGTTTCAGCAAAACTAGTTTCAAAACCAATTTGAGCCAAACTGTAGATGAAGAGGGCTCTGA	3544
Qy	1894	---ACTAAACATATGTTTCATTGACAGACATGCTCCGACCAAGAGCAAGTGGGAACCT	1949
Db	3545	TCTAACAAAATAAGGTTATATGAGTAGACTCTTCAGACCAAGAGCAGCTGGGAATCTGA	3604
Qy	1950	CAGATGCTGC-ACTGTGGAACGCTCTTTTGAAGGGCTGCTGTGCCCTCAAGCC	2004
Db	3605	CATAGGCTTCAATTTGGTGAATCTCTTTAAACAAGGGCTGCAATGCTCATATCCC	3660
RESULT 15			
LOCUS	E15725	2100 bp DNA linear	PAT 28-JUL-1999
DEFINITION	Human mRNA for alpha-1,6-fucosyltransferase, complete cds.		
ACCESSION	E15725		
VERSION	E15725.1	GI:5710498	
KEYWORDS	JP 1998084975-A/1.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 2100)		
TITLE	Taniguchi, N., Uozumi, H. and Yanagiya, S.		
JOURNAL	ALPHA-1,6-FUCOSYLTRANSFERASE GENE DERIVED FROM HUMAN		
COMMENT	PATent: JP 1998084975-A 1 07-APR-1998; TOYOBO CO LTD OS Homo sapiens (human) PN JP 1998084975-A/1 PD 07-APR-1998 PF 17-JUN-1997 JP 1997159692 PR 22-JUL-1996 JP 96P 192260 PI TANIGUCHI, NAOYUKI, UOZUMI HISAFUMI, YANAGIYA SHUSAKU PC C12N15/09, C07H21/04, C12N1/21, C12N9/10, (C12N15/09, C12R1.91), PC (C12N1/21, PC C12R1.19), (C12N9/10, C12R1.19); CC strandedness: Double; CC topology: Linear; FH Key Location/Qualifiers FH source 1. .2100 FT		

```
FT /organism='Homo sapiens'
FT /cell_type='stomach cancer cell' FT
FT /cell_lines='MKN45'
FT 198..1925
FT /product='alpha-1,6-fucosyltransferase'.
FEATURES
  source      Location/Qualifiers
  1..2100
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"
BASE COUNT  651 a  436 c  489 g  524 t
ORIGIN
Query Match      78.3%; Score 1572.8; DB 6; Length 2100;
Best Local Similarity 90.2%; Pred. No. 0;
Matches 1694; Conservative 0; Mismatches 182; Indels 2; Gaps 1;
QY 1 AACAGAACTTATTTTCCTGTGTGCTAACTAGAACCCAGAGTACAAATGTTTCCAAATTCCT 60
DB 97 AACAGAAAGTCTATTACCTGTGCTCACTAACTAGAAACACAGAGTTACAAATGTTTTCAAATTCCT 156
QY 61 TGAGCTCCGAGAGACA--GAAGGGAGTTGAACTCTGAAATGCGGCATGGAGCTGTT 118
DB 157 TGAGCTCCAGGACTCCAGGGAAGTGAAGTTGAAATCTGAAATGCGGCATGGAGCTGTT 216
QY 119 CTGGCGTTGGATTATGCTCATCTTTTTCCTGGGGACCTTATGTTTATATAGGTG 178
DB 217 CTGGCGTTGGATTATGCTCATCTTTTTCCTGGGGACCTTGTCTGTTTATATAGGTG 276
QY 179 GTCAATTTGGTTCGAGATATGACCACTGACCATTTAGCAGAGAACTCTCCAGATTC 238
DB 277 GTCACTTGTGACGAGATATGACCACTCTGATCACTTAGCCGAGAACTCTCCAGATTC 336
QY 239 TTGCAAGCTGAGCGCTTAAACCAACAAATGAAGACTTGAGGAGAAATGGCTGAGTCTC 298
DB 337 TGGCAAGCTTGAAGCTTAAACCAACAGCAATGAAGACTTGAGGCGAAATGGCCGAATCTC 396
QY 299 TCCGAATACCAAGGCCCTATTGATCAGGGGACAGCTACAGAAAGAGTCCGTGTTTATG 358
DB 397 TCCGGATACCAAGGCCCTATTGATCAGGGGCCAGCTATAGGAAGAGTACGGGTTTATG 456
QY 359 AAGAACAGCTTGTAAAGCCAAAGAACAGATTGAAATTTACAGAAACAAAGCTAGGAATG 418
DB 457 AAGAGCAGCTTGTAAAGCCAAAGAACAGATTGAAATTTACAGAAACAAAGCTAGGAATG 516
QY 419 ATCTGGAAAGCATCATGAAATCTTAAAGGAGGAGATTGAAATGGAGCTAAAGAGCTCT 478
DB 517 GTCTGGGAAGGATCATGAAATCTCTGAGGAGGAGATTGAAATGGAGCTAAAGAGCTCT 576
QY 479 GGTGTTTTCTCAAAAGTCAATTTGAAGAAATTAAGAAATTAAGAAAGAAACGAACTCCAAA 538
DB 577 GGTGTTTTCTCAAGAGTCAATTTGAAGAAATTAAGAAATTAAGAAAGAAATGAACCTCCAAA 636
QY 539 GACATGACAGATGAAATCTTTTGGATTTAGGACATCATGAAAGTCTATGACAGATC 598
DB 637 GACATGACAGATGAAATCTTTTGGATTTAGGACATCATGAAAGTCTATAATGACGGATC 696
QY 599 TATACCTACCTCAGTCAACAGATGAGCAGGAGTGGCGGGAAGAAAGCAAGCAAGATC 658
DB 697 TATACCTACCTCAGTCAACAGATGAGCAGGAGTGGCGGGAAGAAAGCAAGCAAGATC 756
QY 659 TGACAGAGCTGTGTCAGCGGAGATAACATATCTGCAGAAATCCCAAGGACTGCAAGCAAG 718
DB 757 TGACAGAACTGTTTCAGCGGAGATAACATATCTTCAGAAATCCCAAGGACTGCAAGCAAG 816
QY 719 CCAGAAAGCTGTATGTAATATCAACAAAGGCTGTGGCTATGGATGTCAATCTCATATG 778
DB 817 CCAGAAAGCTGTGTATATCAACAAAGGCTGTGGCTATGGCTGTGAGCTCCATCATG 876
QY 779 TGGTTTTACTGCTTATGATGCTTATGACCCAGCAACACTCATCTTGAATCTCAGA 838
DB 877 TGGTCTACTGCTTATGATGCTTATGACCCAGCAACACTCATCTTGAATCTCAGA 936
```

```
QY 839 ATTGGCGCTATGCTACTTGGAGGATGGGAGACTGTGTTTAGACCTGTAACTGAGACATGCA 898
DB 937 ATTGGCGCTATGCTACTTGGTGGATGGGAGACTGTATTTAGGCTGTAACTGAGACATGCA 996
QY 899 CAGACAGCTCTGGCCTCTCCACTGGACACTGGTCAGGTGAAGTCAAGGACAAAAATGTTTC 958
DB 997 CAGACAGATCTGGCATCTCCACTGGACACTGGTCAGGTGAAGTGAAGGACAAAAATGTTTC 1056
QY 959 AAGTGGTCGAGCTCCCAATTTGTAGACAGCCTCCATCTCTGCTCTCTTACTTACCTTTGG 1018
DB 1057 AAGTGGTCGAGCTTCCCATTTGTAGACAGACTCTTATCCCGCTCTCTCCATATTTACCTTTGG 1116
QY 1019 CTGTACAGAAAGACCTTTCAGATCGACTCTCTGAGAGTCCATGGTGATCTCTGCAGTGTGGT 1078
DB 1117 CTGTACAGAAAGACCTTCGAGATCGACTTGTACAGTGCATGGTGACCTGACCTGCTGTGGT 1176
QY 1079 GGGTATCCCAAGTTGTCAAAATCTTGTATCCGTCACAACTTGGCTGCGAAAGGAAATAG 1138
DB 1177 GGGTGTCTCAGTTTGTCAAAATCTTGTATCCGCCACAGCTTGGCTAGAAAAAGAAATAG 1236
QY 1139 AAGAAACACCAAGAACCTTGGCTTCAAACTCCAGATTTATGGAGTCCATGTCAGACGCA 1198
DB 1237 AAGAAACACCAAGAACCTTGGCTTCAAACTCCAGATTTATGGAGTCCATGTCAGACGCA 1296
QY 1199 CTGACAAAGTGGGAAACAGAGCAGCTTCCATCCCATTTGAGGAATACATGCTACAGTTG 1258
DB 1297 CAGACAAAGTGGGAAACAGAGCAGCTTCCATCCCATTTGAAGAGTACATGCTGCTGTGGT 1356
QY 1259 AAGAAACATTTTCAGCTTCTCGAACGAGAAATGAAGTGAATAAAAAAGAGTGTATCTGG 1318
DB 1357 AAGAAACATTTTCAGCTTCTTCACGCGAGAAATGCAAGTGGACAAAAAAGAGTGTATTTGG 1416
QY 1319 CCACGTATGACCCCTTCTTTTAAAGGAGGCAAAAGAACTACTCCAAATTTATGAATTTA 1378
DB 1417 CCACAGATGACCCCTTCTTTTAAAGGAGGCAAAACAAAGTACCCCAATTTATGAATTTA 1476
QY 1379 TTAGTGAATACTCTATTTCTTGGTCACTGGAATACACAAACCGATACACAGAAATTTAC 1438
DB 1477 TTAGTGAATACTCTATTTCTTGGTCACTGGAATGCAAACTCGATACACAGAAATTTAC 1536
QY 1439 TTGGGGCGGTGATCCTCGATATACATTTCTCTCCAGGCTGACTTCTTGTGTGTACTT 1498
DB 1537 TTCTGGAGTGTCTCTGATATACATTTCTCTCTCAGGAGACTTCTCTAGTGTGTACTT 1596
QY 1499 TTTCAATCCAGGCTGTAGGGTTGCTTTATGAAATTCATGCAAAACACTGCTATCTGATGCT 1558
DB 1597 TTTCAATCCAGGCTGTGAGTTGCTTTATGAAATTTATGCAAAACACTACATCTCTGATGCT 1656
QY 1559 CTGCAAACTTCCATTTCTTTAGATGACATCTATTTTGGAGGCAAAATGCCCAACACC 1618
DB 1657 CTGCAAACTTCCATTTCTTTAGATGACATCTATTTTGGGGGCCAGAAATGCCCAACATC 1716
QY 1619 AGATTGAGTTTATCTCTCCCACTGAACTAAAGAGGAAATCCCATGSAACCTGGAG 1678
DB 1717 AAATGTCATTTATGCTCAACACCCGAACTGCAGATGAATTTCCATGGAACCTGGAG 1776
QY 1679 ATATCATTTGTTGGCTGGAAACCATTTGGAATGTTTACTCTAAAGGTGTCAACAGAAAC 1738
DB 1777 ATATCATTTGTTGGCTGGAAATCATTTGGGATGCTTATTTCTAAAGGTGTCAACAGGAAT 1836
QY 1739 TAGAAAAACAGGCTGTACCTTCTCAAAAGTCCAGAGAGATGAGAAACAGTCAAT 1798
DB 1837 TGGGAAGGAGCGGCTATATCTCCCTCTCAAAAGTTCGAGAGAGATAGAAACGCTCAAGT 1896
QY 1799 ACCCTACATATCTTGAAGCTGAAAAATAGAGATGGAGTGTAAAGATTTAAACACAGAAAT 1858
DB 1897 ACCCAATATCTTGAAGCTGAGAAATTAAGCTCAGATGGAGAGATTAACGACCAAACT 1956
QY 1859 TAGTTACAGACCATCTCAG 1876
DB 1957 CAGTTCCAGCAAACTCAG 1974
```

Search completed: February 2, 2004, 16:22:47  
Job time : 7393.83 secs

---

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 2, 2004, 10:17:23 ; Search time 543.385 Seconds  
(without alignments)  
9975.371 Million cell updates/sec

Title: US-09-971-773-1

Perfect score: 2008

Sequence: 1 aacagaaactatttctctg.....gctgtgcctcaagcccatg 2008

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2008	100.0	2008	24	ABK70031 Antibody productio
2	1585.6	79.0	3291	23	ABV22468 Human prostate exp
3	1585.6	79.0	3291	23	ABV28283 Human prostate exp
4	1584	78.9	1728	24	ABK70032 Antibody productio
5	1574.2	78.4	3007	22	AAH33315 Human colon cancer
6	1572.8	78.3	2100	18	AAH76574 Human alpha 1-6 fu
7	1514.8	75.4	1759	24	ABA98809 Alpha1,6-fucosyl t
8	1474.2	73.4	1728	18	AAH76573 Pig alpha 1-6 fuco

9	976.4	48.6	979	21	AAC63891 Chinese hamster FU
10	976.4	48.6	979	24	ABK70036 Antibody productio
11	893.2	44.5	979	21	AAC63892 Rat FUT8 cDNA. Ra
12	893.2	44.5	979	24	ABK70037 Antibody productio
13	881	43.9	1017	22	AAF87952 Human alpha 1-6 fu
14	599.8	29.9	699	22	AAH87953 Human alpha 1-6 fu
15	367.6	18.3	2761	23	ABL04601 Drosophila melanog
16	305.6	15.2	394	20	AAH89049 EST clone CB100.
17	302	15.0	9196	24	ABK70033 Antibody productio
18	285.6	14.2	503	22	AAK11265 Human brain expres
19	285.6	14.2	503	22	AAH42879 Probe #11565 used
20	231	11.5	4682	23	ABL04600 Drosophila melanog
21	218.8	10.9	551	22	AAK11037 Human brain expres
22	213.4	10.6	248	22	AAI42659 Probe #11345 used
23	213.4	10.6	248	22	ABA47883 Human breast cell
24	213.4	10.6	248	22	ABA65776 Human foetal liver
25	213.4	10.6	248	22	AAK32861 Probe #11327 for g
26	213.4	10.6	248	22	AAK14177 Human brain expres
27	213.4	10.6	248	22	AAK25142 Human brain expres
28	213.4	10.6	248	22	AAK39911 Human bone marrow
29	213.4	10.6	248	22	AAI20722 Probe #10655 for g
30	213.4	10.6	248	22	AAI45939 Probe #14625 used
31	213.4	10.6	248	22	AAI57183 Probe #25869 used
32	213.4	10.6	248	22	AAI06420 Human liver single
33	213.4	10.6	248	23	ABG39501 Human genome-deriv
34	213.4	10.6	248	24	ABT14010 Breast cancer mark
35	149	7.4	481	25	ABT21901 Human breast cell
36	136.4	6.8	384	22	ABA42755 Human foetal liver
37	136.4	6.8	384	22	ABA53183 Probe #1421 for ge
38	136.4	6.8	384	22	ABA22955 Human brain expres
39	136.4	6.8	384	22	AAK01429 Human bone marrow
40	136.4	6.8	384	22	AAK26877 Probe #1441 for ge
41	136.4	6.8	384	22	AAI11508 Probe #1471 used t
42	136.4	6.8	384	22	AAI22785 Probe #1413 used t
43	136.4	6.8	384	22	AAI01422 Human liver single
44	136.4	6.8	384	23	ABS26469 Human genome-deriv
45	136.4	6.8	384	24	ABS01477

## ALIGNMENTS

### RESULT 1

ABK70031 ID ABK70031 standard; cDNA; 2008 BP.

XX AC ABK70031;

XX DT 15-JUL-2002 (first entry)

XX DE Antibody production method related cDNA #1.

XX KW Antibody production; cytostatic; immunomodulator; vasotropic; virucide;  
XX KW antibacterial; antiinflammatory; antiallergic; allergy; inflammation;  
XX KW autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO;  
XX KW tumour; circulatory disease; infection; primer; ss.

XX OS Cricetulus griseus.

XX PN WO200231140-A1.

XX PD 18-APR-2002.

XX PF 05-OCT-2001; 2001WO-JP08804.

XX PR 06-OCT-2000; 2000JP-0308526.

XX PA (KYOW ) KYOWA HAKKO KOGYO KK.

XX PI Kanda Y, Satoh M, Nakamura K, Uchida K, Shinkawa T, Yamane N;

XX PI Hosaka E, Yamano K, Yamasaki M, Hanai N;

XX XX WPI; 2002-340182/37.

XX Cells producing antibody compositions including antibody fragments and  
PT fusion proteins with Fc domain of antibody, useful for prevention or  
PT treatment of cancer, immune diseases, circulatory diseases and  
PT infections  
XX  
PS Claim 13; Page 1-3; 314pp; Japanese.  
XX  
CC This invention relates to novel method for antibody production  
CC comprising a Chinese hamster ovarian tissue-originated (CHO) cell  
CC transfected with a gene encoding an antibody molecule for producing a  
CC composition comprising an antibody molecule with an Fc domain bonded  
CC to the N-glycoside linkage complex sugar chain. The produced antibody  
CC compositions are drugs for prevention or treatment of diseases  
CC accompanying tumour, allergy or inflammation, autoimmune diseases,  
CC circulatory diseases, and viral and bacterial infections. The  
CC antibodies can be stably produced using the method of the invention  
CC with high binding activity and potency thus leading to high safety and  
CC reduced side effects when applied alone or in combination with other  
CC drugs for therapy. The present sequence represents a nucleotide  
CC molecule used in the method of the invention.  
XX  
SQ Sequence 2008 BP; 626 A; 425 C; 465 G; 492 T; 0 other;  
Query Match 100.0%; Score 2008; DB 24; Length 2008;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AACGAAACTTATTTTCTGCTGCTTAACCTAGAACCCAGATACAAATGTTTCCAAATCTT 60  
DB 1 AACGAAACTTATTTTCTGCTGCTTAACCTAGAACCCAGATACAAATGTTTCCAAATCTT 60  
QY 61 TGAGCTCCGAGAGACAGAGGGAGTGAACCTCTGAAATGCGGGCATGAGCTGTTTCC 120  
DB 61 TGAGCTCCGAGAGACAGAGGGAGTGAACCTCTGAAATGCGGGCATGAGCTGTTTCC 120  
QY 121 TGGCGTTGGATTAATGCTCAATCTTTTCTGCTGGGGACCTTATGTTTATATAGGTGGT 180  
DB 121 TGGCGTTGGATTAATGCTCAATCTTTTCTGCTGGGGACCTTATGTTTATATAGGTGGT 180  
QY 181 CATTTGGTTCGAGATAATGACCACTGACCACTTAGCAGAGAACTCTCAAGATTTCTT 240  
DB 181 CATTTGGTTCGAGATAATGACCACTGACCACTTAGCAGAGAACTCTCAAGATTTCTT 240  
QY 241 GCAAGCTGGAGCGCTTAAACAAACAAATGAAGACTTTGAGAGAAATGGCTGAGTCTCTC 300  
DB 241 GCAAGCTGGAGCGCTTAAACAAACAAATGAAGACTTTGAGAGAAATGGCTGAGTCTCTC 300  
QY 301 CGAATACCAAGAGGCCCTATTGATCAGGGGACAGCTACAGGAAGAGTCCGTGTTTAAAG 360  
DB 301 CGAATACCAAGAGGCCCTATTGATCAGGGGACAGCTACAGGAAGAGTCCGTGTTTAAAG 360  
QY 361 GAACAGCTTTTAAAGCCCAAGACAGATTTGAATTAACGAACCAAGCTTAGGAATGAT 420  
DB 361 GAACAGCTTTTAAAGCCCAAGACAGATTTGAATTTAAGAAACCAAGCTTAGGAATGAT 420  
QY 421 CTGGGAAAGGATCATGAATCTTAAAGGAGGAGTGAATTAAGAGAGCTCTGG 480  
DB 421 CTGGGAAAGGATCATGAATCTTAAAGGAGGAGTGAATTAAGAGCTTAGAGAGCTCTGG 480  
QY 481 TTTTCTTCAAGAGTGAATTAAGAAATTAAGAAATTAAGAGGAAACGAATCTCAAGA 540  
DB 481 TTTTCTTCAAGAGTGAATTAAGAAATTAAGAAATTAAGAGGAAACGAATCTCAAGA 540  
QY 541 CATGAGATGAATTTCTTTGATTTAGGACATGAAAGTCTTATGACAGATCTA 600  
DB 541 CATGAGATGAATTTCTTTGATTTAGGACATGAAAGTCTTATGACAGATCTA 600  
QY 601 TACTACCTCAGTCAACAGATGGAGCGTGTGCTGGGAAAGAAAGAGCAAGATCTG 660  
DB 601 TACTACCTCAGTCAACAGATGGAGCGTGTGCTGGGAAAGAAAGAGCAAGATCTG 660  
QY 661 ACAGAGCTGGTCCAGGGGAGATAACATATCTGCAAGAAATCCCAAGGACTGCGCAAGGCC 720

Db 661 ACAGAGCTGGTCCAGGGGAGATAACATATCTGCAAGAAATCCCAAGGACTGCGCAAGGCC 720  
QY 721 AGAAAGCTGGTATGTAATATCAACAAAGGCTGTGGCTATGGATGTCAATCCATCATGTG 780  
DB 721 AGAAAGCTGGTATGTAATATCAACAAAGGCTGTGGCTATGGATGTCAATCCATCATGTG 780  
QY 781 GTTTACTGCTTATGATTTGCTTATGGCAACCAGCGAACACTCATCTTTGGAATCTCAGAA 840  
DB 781 GTTTACTGCTTATGATTTGCTTATGGCAACCAGCGAACACTCATCTTTGGAATCTCAGAA 840  
QY 841 TGGCGCTATGCTACTGAGGATGGAGACTGTGTTTAGACCTGTAAGTGAACATGACACA 900  
DB 841 TGGCGCTATGCTACTGAGGATGGAGACTGTGTTTAGACCTGTAAGTGAACATGACACA 900  
QY 901 GACAGGCTGCGCTCTCCACTGGACACTGGTCAAGTGAAGTGAAGGACAAAGAAATGTTCAA 960  
DB 901 GACAGGCTGCGCTCTCCACTGGACACTGGTCAAGTGAAGTGAAGGACAAAGAAATGTTCAA 960  
QY 961 GTGCTCGAGCTCCCAATTTAGACAGCTCCATCTGCTCTCTTACTTACCTTGGCT 1020  
DB 961 GTGCTCGAGCTCCCAATTTAGACAGCTCCATCTGCTCTCTTACTTACCTTGGCT 1020  
QY 1021 GTACAGAGACCTTGGAGACTCGATCGATGATCGATGATCGATGATCGATGATGATG 1080  
DB 1021 GTACAGAGACCTTGGAGACTCGATCGATGATCGATGATCGATGATGATGATGATG 1080  
QY 1081 GTATCCAGTTTGTCAAAATACTTCCATCCGTCACAACTTGGCTGGAAAGGAAATAGAA 1140  
DB 1081 GTATCCAGTTTGTCAAAATACTTCCATCCGTCACAACTTGGCTGGAAAGGAAATAGAA 1140  
QY 1141 GAAACCAAGAGCTTGGCTTCAAACTCCAGTATTTGGAGTCCATGTGAGAGGCACT 1200  
DB 1141 GAAACCAAGAGCTTGGCTTCAAACTCCAGTATTTGGAGTCCATGTGAGAGGCACT 1200  
QY 1201 GACAAAGTGGGACAGAGAGCGCTTCCATCCCAATGAGGAATACATGTTACAGTTGAA 1260  
DB 1201 GACAAAGTGGGACAGAGAGCGCTTCCATCCCAATGAGGAATACATGTTACAGTTGAA 1260  
QY 1261 GAACATTTTCAAGCTTCTCGAACGAGAAATGAAAGTGAATGAAAGAGTGTATCTGACC 1320  
DB 1261 GAACATTTTCAAGCTTCTCGAACGAGAAATGAAAGTGAATGAAAGAGTGTATCTGACC 1320  
QY 1321 ACTGATGACCTTCTTTTAAAGGAGGCAAGACAAAGTACTCAATTTATGAATTTAT 1380  
DB 1321 ACTGATGACCTTCTTTTAAAGGAGGCAAGACAAAGTACTCAATTTATGAATTTAT 1380  
QY 1381 AGTGATTAATCTATTTCTTGGTCAAGTCACTACACACGATACACAGAAATTCACCT 1440  
DB 1381 AGTGATTAATCTATTTCTTGGTCAAGTCACTACACACGATACACAGAAATTCACCT 1440  
QY 1441 CGGGCGTGATCTCGGATATACACTTCTCTCCAGGCTGACTTCTTCTGTGTACTTTT 1500  
DB 1441 CGGGCGTGATCTCGGATATACACTTCTCTCCAGGCTGACTTCTTCTGTGTACTTTT 1500  
QY 1501 TCATCCCAAGGCTGTAGGGTGTGTAATGAAATCATGCAAAACACTGCACTCTGATGCTCT 1560  
DB 1501 TCATCCCAAGGCTGTAGGGTGTGTAATGAAATCATGCAAAACACTGCACTCTGATGCTCT 1560  
QY 1561 GCAAACTTCAATCTTTAGATGACATCTATTTTGGAGGCCAAATGCCCCCAACCCAG 1620  
DB 1561 GCAAACTTCAATCTTTAGATGACATCTATTTTGGAGGCCAAATGCCCCCAACCCAG 1620  
QY 1621 ATTGAGTTTATCTCTACCAACTCGAACTTAAAGAGGAAATCCCATGGAACCTGGAGAT 1680  
DB 1621 ATTGAGTTTATCTCTACCAACTCGAACTTAAAGAGGAAATCCCATGGAACCTGGAGAT 1680  
QY 1681 ATCAATGGTGTGGTGAACCAATTTGGAATGGTTTACTTAAAGGTGTCAACAGAAACTA 1740  
DB 1681 ATCAATGGTGTGGTGAACCAATTTGGAATGGTTTACTTAAAGGTGTCAACAGAAACTA 1740  
QY 1741 GGAAGAACGCGCTGATCTCTCTTACAAAGTCCGAGAGAGATAGAAACAGTCAAAATAC 1800

Db 1741 GGAACAAACAGCGCTGTACCTCTTCTACAAAGTCCGAGAGAGATAGAAACAGTCAAATAC 1800  
Qy 1801 CCTACATATCTGAAGCTGAAATAATAGAGATGGAGTGAAGAGATTAACAACAGAAATTA 1860  
Db 1801 CCTACATATCTGAAGCTGAAATAATAGAGATGGAGTGAAGAGATTAACAACAGAAATTA 1860  
Qy 1861 GTTCAGACCATCTCAGCCCAAGCAGAGAACCCAGACATAACATATGTTTCATTGACAGACAT 1920  
Db 1861 GTTCAGACCATCTCAGCCCAAGCAGAGAACCCAGACATAACATATGTTTCATTGACAGACAT 1920  
Qy 1921 GCTCGCACCAAGAGCAAGTGGGAACCCCTCAGATGCTGCACTGGTGGGAACCGCTTTTGT 1980  
Db 1921 GCTCGCACCAAGAGCAAGTGGGAACCCCTCAGATGCTGCACTGGTGGGAACCGCTTTTGT 1980  
Qy 1981 GAAGGGCTGCTGTCCTCCAGCCCATG 2008  
Db 1981 GAAGGGCTGCTGTCCTCCAGCCCATG 2008

## RESULT 2

ABV22468

ID ABV22468 standard; cDNA; 3291 BP.

XX AC

XX ABV22468;

XX DT 13-SEP-2002 (first entry)

XX DE

XX Human prostate expression marker cDNA 22459.

XX KW

XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
XX pharmacogenomic marker; gene; ss.

XX OS

XX Homo sapiens.

XX PN

XX WO200160860-A2.

XX PD

XX 23-AUG-2001.

XX PF

XX 20-FEB-2001; 2001WO-US05171.

XX PR

XX 17-FEB-2000; 2000US-183319P.

XX PR

XX 16-MAR-2000; 2000US-189862P.

XX PR

XX 25-MAY-2000; 2000US-207454P.

XX PR

XX 09-JUN-2000; 2000US-211314P.

XX PR

XX 18-JUL-2000; 2000US-219007P.

XX PR

XX 13-DEC-2000; 2000US-255281P.

XX PA

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI

XX Schlegel R, Endege WO, Monahan JB;

XX XX

XX WPI; 2001-662795/76.

XX DR

XX Novel isolated nucleic acid molecule associated with cancerous state of

XX PT

XX prostate cells and correlating with presence of prostate cancer, useful  
XX for detecting presence of prostate cancer, stage of prostate cancer -

XX PS

XX Claim 1; Page 3914; 11750pp; English.

XX CC

XX The invention relates to an isolated nucleic acid molecule (I) comprising  
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the

XX CC

XX specification or its complement. (I) is useful for:

XX CC

XX (a) assessing whether a patient is afflicted with prostate cancer;

XX CC

XX (b) monitoring the progression of prostate cancer in a patient;

XX CC

XX (c) assessing the efficacy of a test compound to inhibit prostate

XX CC

XX cancer in a patient;

XX CC

XX (d) assessing the efficacy of a therapy for inhibiting prostate cancer

XX CC

XX in a patient;

XX CC

XX (e) selecting a composition for inhibiting prostate cancer in a patient;

XX CC

XX (f) assessing the prostate cell carcinogenic potential of a compound;

XX CC

XX (g) determining whether prostate cancer has metastasized in a patient;

XX CC

XX (h) assessing the aggressiveness or indolence of prostate cancer in a

XX CC

XX patient;

XX CC

CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX

SQ Sequence 3291 BP; 925 A; 729 C; 745 G; 871 T; 21 other;

XX

Query Match 79.0%; Score 1585.6; DB 23; Length 3291;

XX

Best Local Similarity 87.7%; Pred. No. 0;

XX

Matches 1789; Conservative 0; Mismatches 219; Indels 32; Gaps 4;

XX

Qy 1 AACGAAAGCTTATTTCTGTGGCTAACTAGAACACAGAGTACAATGTTTCCAAATCTT 60

XX

Db 626 AACGAAAGCTTATTTCTGTGGCTAACTAGAACACAGAGTACAATGTTTCCAAATCTT 685

XX

Qy 61 TGAGCTCCGAGAAACACA--GAAGGAGTGAAGAACTCTGAAATCGGGCATGAGCTGTT 118

XX

Db 686 TGAGCTCCGAGAAACACA--GAAGGAGTGAAGAACTCTGAAATCGGGCATGAGCTGTT 745

XX

Qy 119 CCTGGCTGGATATGCTCATTTCTTTGCTGGGGACCTTATTTGTTTATATAGTGT 178

XX

Db 746 CCTGGCTGGATATGCTCATTTCTTTGCTGGGGACCTTATTTGTTTATATAGTGT 805

XX

Qy 179 GTCAATTTGTTTCGAGATAATGACCACCTGACCATTTCTAGCAGAGAACTCTCCAAGATTC 238

XX

Db 806 GTCAATTTGTTTCGAGATAATGACCACCTGACCATTTCTAGCAGAGAACTCTCCAAGATTC 865

XX

Qy 239 TTGCAAGCTGGAGCGCTTAAACAAACAAATGAAGCTTGAAGGAGAAATGGCTGAGTCTC 298

XX

Db 866 TGCAAGCTGGAGCGCTTAAACAAACAAATGAAGCTTGAAGGAGAAATGGCTGAGTCTC 925

XX

Qy 299 TCCGAATACAGAAAGCCCTTATTCATCGGGGACAGCTACAGGAAGAGTCTGTTTGTAG 358

XX

Db 926 TCCGAATACAGAAAGCCCTTATTCATCGGGGACAGCTACAGGAAGAGTCTGTTTGTAG 985

XX

Qy 359 AAGAACAGCTTGTAAAGGCCAAAGAACAGATTGAATAATTACAAGAAACAACTAGGAATG 418

XX

Db 986 AAGAACAGCTTGTAAAGGCCAAAGAACAGATTGAATAATTACAAGAAACAACTAGGAATG 1045

XX

Qy 419 ATCTGGGAAAGGATCATGAAATCTTAAAGGAGGAGGATTGAAATATGGAGCTTAAAGAGCTCT 478

XX

Db 1046 ATCTGGGAAAGGATCATGAAATCTTAAAGGAGGAGGATTGAAATATGGAGCTTAAAGAGCTCT 1105

XX

Qy 479 GGTCTTTCTACAAAGTGAATTTGAAGAAATTAAGAAATTAAGAGGAACCAATCTCCAA 538

XX

Db 1106 GGTCTTTCTACAAAGTGAATTTGAAGAAATTAAGAGGAACCAATCTCCAA 1165

XX

Qy 539 GACATGCAGATGAATTTCTTTTGGATTTAGACATCATGAAAGTCTCATCATGACAGATC 598

XX

Db 1166 GACATGCAGATGAATTTCTTTTGGATTTAGACATCATGAAAGTCTCATCATGACAGATC 1225

XX

Qy 599 TATACTACCTCAGTCAAAACAGATGGAGCGAGTGGCGGGAAGGAAAGGAAAGGAAAGATC 658

XX

Db 1226 TATACTACCTCAGTCAAAACAGATGGAGCGAGTGGCGGGAAGGAAAGGAAAGGAAAGATC 1285

XX

Qy 659 TGACAGAGCTGGTCCAGCGGAGATTAACATATCTGAGAAATCCAGAGCTGACAGCAAG 718

XX

Db 1286 TGACAGAGCTGGTCCAGCGGAGATTAACATATCTGAGAAATCCAGAGCTGACAGCAAG 1345

XX

Qy 719 CCAGAAAGCTGGTATGTAATATCAAAAGGCTGGGCTATGGATGTCAATCCCATCATG 778

XX

Db 1346 CCAGAAAGCTGGTATGTAATATCAAAAGGCTGGGCTATGGATGTCAATCCCATCATG 1405

XX

Qy 779 TGGTTACTGCTCATCATATTCCTTATGGCAACCCAGCAACACTCATCTTGAATCTCAGA 838

XX

Db 1406 TGGTTACTGCTCATCATATTCCTTATGGCAACCCAGCAACACTCATCTTGAATCTCAGA 1465

XX

Qy 839 ATTGCGGCTATGCTACTGGAGGATGGGAGACTGTTTGTAGACCTGTAGTGAAGACATGCA 898

XX

Db 1466 ATTGCGGCTATGCTACTGGAGGATGGGAGACTGTTTGTAGACCTGTAGTGAAGACATGCA 1525

XX

Qy 899 CAGACAGCTGGGCTCTCCACTGGACACTGGTCAAGTGAAGTGAAGCAAAAATGTTTC 958

XX

Db 1526 CAGACAGCTGGGCTCTCCACTGGACACTGGTCAAGTGAAGTGAAGCAAAAATGTTTC 1585

XX

Qy 959 AAGTGGTCGAGCTCCCCATTGTAGACAGCTCCATCCTCGTCTCTTACTTACCCTTGG 1018

XX





```
PF 05-OCT-2001; 2001WO-JP08804.
XX
PR 06-OCT-2000; 2000JP-0308526.
XX
XX (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI Kanda Y, Satoh M, Nakamura K, Uchida K, Shinkawa T, Yamane N;
PI Hosaka E, Yamano K, Yamasaki M, Hanai N;
XX
XX WPI; 2002-340182/37.
XX
PT Cells producing antibody compositions including antibody fragments and
PT fusion proteins with Fc domain of antibody, useful for prevention or
PT treatment of cancer, immune diseases, circulatory diseases and
PT infections
XX
XX Claim 32; Page 4-11; 314pp; Japanese.
XX
CC This invention relates to novel method for antibody production
CC comprising a Chinese hamster ovarian tissue-originated (CHO) cell
CC transfected with a gene encoding an antibody molecule for producing a
CC composition comprising an antibody molecule with an Fc domain bonded
CC to the N-glycoside linkage complex sugar chain. The produced antibody
CC compositions are drugs for prevention or treatment of diseases
CC accompanying tumour, allergy or inflammation, autoimmune diseases,
CC circulatory diseases, and viral and bacterial infections. The
CC antibodies can be stably produced using the method of the invention
CC with high binding activity and potency thus leading to high safety and
CC reduced side effects when applied alone or in combination with other
CC drugs for therapy. The present sequence represents a nucleotide
CC molecule used in the method of the invention.
XX
SQ Sequence 1728 BP; 529 A; 365 C; 400 G; 434 T; 0 other;

Query Match 78.9%; Score 1584; DB 24; Length 1728;
Best Local Similarity 94.8%; Pred. No. 0;
Matches 1638; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 100 ATCGGGGATGACTGGTTCCTGGGCTTGGATTGCTCAATCTTTTGGCTGGGGGACC 159
DB 1 ATCGGGGATGACTGGTTCCTGGGCTTGGATTGCTCAATCTTTTGGCTGGGGGACC 60

QY 160 TTATGTTTTATATAGTGGTTCATTTGGTTCGAGATAATGACCCCTGACCATCTTAGC 219
DB 61 TTGTTATTTTATATAGTGGTTCATTTGGTTCGAGATAATGACCCCTGATCACTCCAGC 120

QY 220 AGAGAACTCTCAAGATCTTTCGAAGCTGGAGCGCTTAAACACAAATGAAGACTTG 279
DB 121 AGAGAACTCTCAAGATCTTTCGAAGCTTGAACGCTTAAACACAAATGAAGACTTG 180

QY 280 AGGAGAACTGGCTGAGTCTCTCGAATACCAAGAGCCCTATTGATCAGGGGACAGCTACA 339
DB 181 AGGCGAATGGCTGAGTCTCTCGAATACCAAGAGCCCTATTGATCAGGGGACAGCTACA 240

QY 340 GGAAGAGTCCGTTGTTTTAGAGAACAGCTTGTAAAGGCCAAAGAACAGATGAAATTAAC 399
DB 241 GGAAGAGTCCGTTGTTTTAGAGAACAGCTTGTAAAGGCCAAAGAACAGATGAAATTAAC 300

QY 400 AAGAAACAGCTAGGAATGATCTGGGAAGGATCATGAAATCTTAAGGAGGAGGATGAA 459
DB 301 AAGAAACAGCTAGGAATGATCTGGGAAGGATCATGAAATCTTAAGGAGGAGGATGAA 360

QY 460 AATGAGCTTAAAGAGCTCTGGTTTTCTTACAAAGTGAATGAAGAAATTAAGAAATTA 519
DB 361 AATGAGCTTAAAGAGCTCTGGTTTTCTTACAAAGGAACTGAAGAAATTAAGAAATTA 420

QY 520 GAAGGAAACGAATCCAAAGACATCCAGATGAAATCTTTGGATTTAGGACATCATGAA 579
DB 421 GAAGGAAATGAATCCAAAGACATCCAGATGAAATCTTTGGATTTAGGACATCATGAA 480

QY 580 AGGCTTATCATGACAGATCTATCTACTCTCAGTCAACAGATGGAGCAGGTGAGTGGGG 639
DB 481 AGGCTTATCATGACAGATCTATCTACTCTCAGTCAACAGATGGAGCAGGTGAGTGGGG 540
```

---

```
QY 640 GAAAAGAGGCCAAAGATCTGACAGAGCTGGTCCAGCGGAGAAATAACATATCTCGAGAA 699
DB 541 GAAAAGAGGCCAAAGATCTGACAGAGCTGGTCCAGCGGAGAAATAACATATCTCGAGAA 600

QY 700 CCCAAGGACTGCAGCAAAAGCCAGAAAGCTGTATGTAATATCAACAAAGGCTGTGGCTAT 759
DB 601 CCTAAGGACTGCAGCAAAAGCCAGAAAGCTGTGTGTAACATCAATATAAAGGCTGTGGCTAT 660

QY 760 GGATGTCAATCTCAATCATGTGCTTCTATGATTTGCTTATGGCACCACCGAACA 819
DB 661 GGTGTCAATCTCAATCATGTGCTTCTATGATTTGCTTATGGCACCACCGAACA 720

QY 820 CTCAATCTTGAATCTCAAGATTTGGCGCTATGCTACTGAGGATGGGAGATGCTGTGTA 879
DB 721 CTCAATCTTGAATCTCAAGATTTGGCGCTATGCTACTGAGGATGGGAGATGCTGTGTA 780

QY 880 CCTGTAGTGAGCATGCACAGACAGCTCTGGCCCTCTCCACTGGACACTGGTCAGGTGAA 939
DB 781 CCTGTAGTGAGCATGCACAGACAGCTCTGGCCCTCTCCACTGGACACTGGTCAGGTGAA 840

QY 940 GTGAAGGACAAAATGTTCAAGTGGTCTGAGCTCCCAATTTGTAGACAGCTCCATCTCGT 999
DB 841 GTAAATGACAAAACATTCAGTGGTCTGAGCTCCCAATTTGTAGACAGCTCCATCTCGG 900

QY 1000 CCTCTTACTTACCCCTTGGCTGTACAGAAAGACCTTGCAGATCGACTCTGAGAGTCCAT 1059
DB 901 CCTCTTACTTACCCCTTGGCTGTTCAGAAAGACCTTGCAGACCCGACTCTTAAGAGTCCAT 960

QY 1060 GGTGATCTCTCAGTGTGGTGTATCCAGTGTTCCTCAATCTGTAATCTGATCGTCCCAACCT 1119
DB 961 GGTGATCTCTCAGTGTGGTGTATCCAGTGTTCCTCAATCTGTAATCTGATCGTCCCAACCT 1020

QY 1120 TGGCTGGAAGGGAAATAGAAAGAACCAACCAAGAGCTTGGCTTCAACATCTCCAGTTAT 1179
DB 1021 TGGCTGGAAGGGAAATAGAAAGAACCAACCAAGAGCTTGGCTTCAACATCTCCAGTTAT 1080

QY 1180 GGAGTCCATGTACAGCCGACTGACAAAGTGGGAAACAGAGCAGCTTCCATCCCATGAG 1239
DB 1081 GGAGTCCATGTACAGCCGACTGACAAAGTGGGAAACAGAGCAGCTTCCATCCCATGAG 1140

QY 1240 GAATACATGTTACAGCTTGAAGACATTTTCAGCTTCTCGAACGACAGATGAAAGTGGAT 1299
DB 1141 GAGTACATGTTACAGCTTGAAGACATTTTCAGCTTCTCGAACGACAGATGAAAGTGGAT 1200

QY 1300 AAAAAAGAGTGTATCTGGCCACTGATGACCCCTTCTTTGTTTAAAGAGGCAAGACAAAG 1359
DB 1201 AAAAAAGAGTGTATCTGGCTACTGATGATCTTCTTTTAAAGAGGCAAGACAAAG 1260

QY 1360 TACTCCAAATTAAGATTTATAGTGATACTCTATTTCTTGGTTCAGCTGGGACTACAAAC 1419
DB 1261 TACTCCAAATTAAGATTTATAGTGATACTCTATTTCTTGGTTCAGCTGGGACTACAAAC 1320

QY 1420 CGATACAGAAAATTCACCTTCGGGGGTGATCTGGATATACACTTCTCTCCAGGCT 1479
DB 1321 CGGTACAGAAAATTCACCTTCGGGGGTGATCTGGATATACACTTCTCTCCAGGCT 1380

QY 1480 GACTTCTTGTGTACTTTTCAATCCAGGTCTGTAGGGTGTGCTTATGAAATCATGCAA 1539
DB 1381 GACTTCTTGTGTACTTTTCAATCCAGGTCTGTAGGGTGTGCTTATGAAATCATGCAA 1440

QY 1540 AACTGTCATCTGATGCTCTGAAAATTCATTTCTTTAGATGATCATCTACTATTTTGA 1599
DB 1441 ACCCTGTCATCTGATGCTCTGAAAATTCATTTCTTTAGATGATCATCTACTATTTTGA 1500

QY 1600 GGCCAAAATGCCCAACACAGATTTGAGTTTATCTCCACCAACCTCCGAACTAAGAGGAA 1659
DB 1501 GGCCAAAATGCCCAACACAGATTTGAGTTTATCTCCACCAACCTCCGAACTAAGAGGAA 1560

QY 1660 ATCCCAATGGAACCTGGAGATATCATTTGGTGTGGCTGGAAACCAATTTGGAATGTTACTCT 1719
DB 1561 ATCCCAATGGAACCTGGAGATATCATTTGGTGTGGCTGGAAACCAATTTGGATGTTTCT 1620
```





```

QY 1138 GAAGAAACACCAAGAAGCTTGGCTTCAAAATCCAGTATTGGAGTCCATGTTCAGACGC 1197
  |||
  |||
  |||
Db 1459 GAAGAAGCCACCAAGAAGCTTGGCTTCAAAATCCAGTATTGGAGTCCATGTTCAGACGC 1518
  |||
  |||
  |||
QY 1198 ACTGACAAAGTGGGAACAGAGCCTTCCATCCATTTGAGGAATACATGTTACAGCTT 1257
  |||
  |||
  |||
Db 1519 ACAGACAAAGTGGGAACAGAGCCTTCCATCCATTTGAGGAATACATGTTACAGCTT 1578
  |||
  |||
  |||
QY 1258 GAAGAACATTTTCAGCTTCTCGAACCGAGAAATGAAGTGGATAAAAAAGAGTGTATCTG 1317
  |||
  |||
  |||
Db 1579 GAAGAACATTTTCAGCTTCTCGAACCGAGAAATGAAGTGGATAAAAAAGAGTGTATTTG 1638
  |||
  |||
  |||
QY 1318 GCCACTGATGACCCCTTCTTTGTTAAAGGAGGCAAAAGACTCTCCAAATTTATGAATTT 1377
  |||
  |||
  |||
Db 1639 GCCACAGATGACCCCTTCTTTAATAAGGAGGCAAAACAAAGTACCCCAATTTATGAATTT 1698
  |||
  |||
  |||
QY 1378 ATTAGTGATACTCTATTTCTTGGTGCAGCTGGACTACACAACCGATACACAGAAATTTCA 1437
  |||
  |||
  |||
Db 1699 ATTAGTGATACTCTATTTCTTGGTGCAGCTGGACTACACAACCGATACACAGAAATTTCA 1758
  |||
  |||
  |||
QY 1438 CTTCCGGGGTGATCTCGATATACATTTCTCTCCAGGCTGACTTCTTGTGTGTACT 1497
  |||
  |||
  |||
Db 1759 CTTCTGGAGTGATCTGGATATACATTTCTCTCTCAGGCGAGACTTCTTGTGTGTACT 1818
  |||
  |||
  |||
QY 1498 TTTTCATCCAGGTCTGTAGGTTGCTTATGAATCATGCAAACTGCATCTCGATGCC 1557
  |||
  |||
  |||
Db 1819 TTTTCATCCAGGTCTGTAGGTTGCTTATGAATCATGCAAACTGCATCTCGATGCC 1878
  |||
  |||
  |||
QY 1558 TCTGAAATCTCATTTTATAGATGACATCTACTATTTTGGAGGCAAAATGCCACAAAC 1617
  |||
  |||
  |||
Db 1879 TCTGAAATCTCATTTTATAGATGACATCTACTATTTTGGAGGCAAAATGCCACAAAC 1938
  |||
  |||
  |||
QY 1618 CAGATTGAGTTTATCTCACCACCTCGAATCTGAAGGAATCCCATGGAACCTGGA 1677
  |||
  |||
  |||
Db 1939 CAAATGGCAATTTATGCTCACCACCCGAACTGCAGATGAATTTCCCATGGAACCTGGA 1998
  |||
  |||
  |||
QY 1678 GATATCATTTGGTGGCTGGAAACCAATTTGGAATGGTTACTCTTAAAGGTGTCAACAGAAA 1737
  |||
  |||
  |||
Db 1999 GATATCATTTGGTGGCTGGAAATCATTTGGGATGGCTATTCTTAAAGGTGTCAACAGAAA 2058
  |||
  |||
  |||
QY 1738 CTAGAAACAAAGGCTGTACCTTCTTACAAAGTCCGAGAGAAATAGAAACAGTCAAA 1797
  |||
  |||
  |||
Db 2059 TTGGGAAAGGACGGGCTATATCTCTTACAAAGTTCGAGAGAAATAGAAACCGTCAAG 2118
  |||
  |||
  |||
QY 1798 TACCCTACATATCTCGAAGCTGAAATAGAGATGAGTGTGAAGAGATTAA----- 1848
  |||
  |||
  |||
Db 2119 TACCCTACATATCTCGAAGCTGAAATAGAGATGAGTGTGAAGAGATTAAACGACCAAC 2178
  |||
  |||
  |||
QY 1849 -----CAACAGAAATTTAGTTTCCAGACCATCTCAGCCCAAGCAGAACCCAG----- 1893
  |||
  |||
  |||
Db 2179 TCAGTTCCGACCAACTCAGTTCAACACCATTTCCAGCCCAACTGTAGTGAAGAGGCTCTG 2238
  |||
  |||
  |||
QY 1894 -----ACTAACATATGTTTCATTTGACAGATGCTCCGACCAAGAGCAAGTGGAAACC 1948
  |||
  |||
  |||
Db 2239 ATCTAACAAATAAGGTTATATGAGTAGATCTCTCAGCACCACAGAGAGTGGGAATG 2298
  |||
  |||
  |||
QY 1949 TCAGATGCTGC-ACGTGGTGGAGCGCTCTTTGTGGAAGGCTGCTGTGCCCTCAAGCCCAT 2007
  |||
  |||
  |||
Db 2299 ACATAGGCTTCAATTTGGTGAATTTCTCTTTAAAGAGGCTGCAATGCCCTCATACCCAT 2358
  |||
  |||
  |||
QY 2008 G 2008
  |||
  |||
  |||
Db 2359 G 2359
  |||
  |||
  |||

```

## RESULT 6

AAT76574

ID AAT76574 standard; cDNA to mRNA; 2100 BP.

XX

AC AAT76574;

XX

DT 05-MAR-1998 (first entry)

```

XX Human alpha 1-6 fucosyltransferase gene.
DE
XX Alpha 1-6 fucosyltransferase; enzyme; pig; human; fucose transfer;
KW guanosine diphosphate; sugar chain synthesis; modification; antibody;
KW GlcNAc; cancer diagnosis; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
FT CDS 198..1925
FT /*tag= a
XX WO9727303-A1.
XX 31-JUL-1997.
XX 23-JAN-1997; 97WO-JP00171.
XX 22-JUL-1996; 96JP-0192260.
XX 24-JAN-1996; 96JP-0010365.
XX 21-JUN-1996; 96JP-0161648.
XX 24-JUN-1996; 96JP-0162813.
XX (TOYM ) TOYO BOSEKI KK.
XX Shiba T, Taniguchi N, Uozumi N, Yanagidani S;
XX WPI; 1997-393690/36.
XX P-PSDB; AAW22125.
XX Human or pig alpha 1-6 fucosyltransferase and DNA encoding it - for
XX synthesis and modification of sugar chains and used as an antigen
XX for production of diagnostic antibodies
XX Claim 18; Page 39-43; 61pp; Japanese.
XX AAT76573 and AAT76574 represent the coding sequences for the pig and
XX human alpha 1-6 fucosyltransferases of the invention, respectively. The
XX encoded enzyme transfers fucose from guanosine diphosphate to the
XX 6-hydroxyl group of the GlcNAc nearest to R in the receptor molecule:
XX (GlcNAc)beta 1-2Manalpha 1-6) (GlcNAc)beta 1-2Manalpha 1-3)Manbeta
XX 1-4GlcNAc)beta 1-4GlcNAc)beta 1-4GlcNAc)beta 1-6) (Fucalpha 1-6) (GlcNAc)beta
XX 1-2Manalpha 1-3)Manbeta 1-4GlcNAc)beta 1-4GlcNAc)beta 1-4GlcNAc)beta
XX an optimum pH of about 7.0 (pig) or 7.5 (human), and is stable over the
XX pH range 4-10 after 5 hours at 4 degrees C. The optimum working
XX temperature of the enzyme is 30-37 degrees C. A bivalent metal is not
XX required for activity of the enzyme, and the enzyme is not inhibited in
XX the presence of 5 mM EDTA. The enzyme is useful in the synthesis and
XX modification of sugar chains, and as antigen for the production of
XX antibodies recognising the enzyme. The antibodies can be used for the
XX diagnosis of cancer and other diseases.
XX Sequence 2100 BP; 551 A; 436 C; 489 G; 524 T; 0 other;

```

```

Query Match 78.3%; Score 1572.8; DB 18; Length 2100;
Best Local Similarity 90.2%; Pred. No. 0;
Matches 1694; Conservative 0; Mismatches 182; Indels 2; Gaps 1;

```

```

QY 1 AACGAAACTTATTTTCTGTGTGCTTAACATAGAACACAGAGTACAATGTTTCCAAATCTT 60
  |||
  |||
  |||
Db 97 AACGAAAGTCTATTTTCACTGTGCACTAACTAGAACACAGAGTACAATGTTTCCAAATCTT 156
  |||
  |||
  |||
QY 61 TGAGCTCCGAGAGACA--GAAGGAGTGTGAACCTCTGAAATCGGGCATGGACTGGTT 118
  |||
  |||
  |||
Db 157 TGAGCTCCAGGACTCCAGGGAAGTGAATCTGAAATCTGAAATCGGCCATGGACTGGTT 216
  |||
  |||
  |||
QY 119 CCTGGCGTTGGATTATGCTCATTTCTTTTGTGCTGGGAGACCTTATGTTTATATAGGTG 178
  |||
  |||
  |||
Db 217 CCTGGCGTTGGATTATGCTCATTTCTTTTGTGCTGGGAGACCTTGTCTGTTTATATAGGTG 276
  |||
  |||
  |||
QY 179 GTCAATTTGTTTCGAGATAATCACCACCTGACCATTTCTAGCAGAGAACTCTCCAAAGATTG 238
  |||
  |||
  |||

```



Db 277 GTCACTTGGTACGAGATAATGACCATCTGATCACTCTAGCGGAACTGTCCAGATTC 336  
Qy 239 TTGCAAGCTGAGCGCTTAAACCAACAAATGAAGACTTGAGGAGATGCTGAGTCTC 298  
Db 337 TGGCAAGCTTGAACGCTTTAAACAGCAGAAATGAAGACTTGAGGCGAATGCGCAATCTC 396  
Qy 299 TCCGAATACCAAGAGCGCTTATGATCAGGGGACAGCTACAGGAGAGTCCGCTGTTTTAG 358  
Db 397 TCCGATACCAAGAGCGCTTATGATCAGGGGCGAGCTATAGGAAGAGTACGCGTTTTAG 456  
Qy 359 AAGAACAGCTGTTTAAAGCCAAAGAAACAGATTTAAATTTACAGAAACAGCTAGGAATG 418  
Db 457 AAGACAGCTGTTTAAAGCCAAAGAAACAGATTTAAATTTACAGAAACAGACCCAGAAATG 516  
Qy 419 ATCTGGGAAAGGATCATGAAATCTTAAAGGAGGAGATTGAAATTTAGGAGCTTAAAGAGCTCT 478  
Db 517 GTCTGGGAAGGATCATGAAATCTTGAAGAGGAGATTGAAATTTAGGAGCTTAAAGAGCTCT 576  
Qy 479 GGTTTTTTCTCAAAAGTGAATTTGAAGAAATTTAAAGAAATTTAGAGGAAACGAACTCCAAA 538  
Db 577 GGTTTTTTCTCAAGAGTGAATTTGAAGAAATTTAAAGAACTTTAGAGGAAATGAATCTCCAAA 636  
Qy 539 GACATGAGATGAATTTCTTTTGGATTTAGGACATCATGAAGTCTATCATGACAGATC 598  
Db 637 GACATGAGATGAATTTCTTTTGGATTTAGGACATCATGAAGTCTATTAATGACGATC 696  
Qy 599 TATACACTCAGTCAAAAGATGAGCAGAGTGTGAGTGGCGGAAAGAAAGCCAAAGATC 658  
Db 697 TATACACTCAGTCAAGATGAGCAGAGTGTGAGTGGCGGAAAGAAAGCCAAAGATC 756  
Qy 659 TGACAGAGCTGTGTCAGCGGAGATAACATATCTCAGAAATCCCAAGGACTGCAGCAAG 718  
Db 757 TGACAGAACTGTGTCAGCGGAGATAACATATCTCAGAAATCCCAAGGACTGCAGCAAG 816  
Qy 719 CCAGAAAGCTGTATGTAATTCACAAAGGCTGTGGCTATGGATGTCAACTCCATCATG 778  
Db 817 CCAAAAAGCTGTGTGTAATTCACAAAGGCTGTGGCTATGGCTGTGAGCTCCATCATG 876  
Qy 779 TGGTTACTGCTTCATGATTTGCTTATGGACCCAGGACACTCATCTTGGATCTCAGA 838  
Db 877 TGGTCTACTGCTTCATGATTTGATATGGACCCAGGACACTCATCTTGGATCTCAGA 936  
Qy 839 ATTGGCGCTATGCTACTGAGATGGGAGACTGTGTTAGACTCTGTAAGTGAACATGCA 898  
Db 937 ATTGGCGCTATGCTACTGAGTGGGAGACTGTATTTAGGCTGTGAAGTGAACATGCA 996  
Qy 899 CAGACAGGCTGGCTCTCCACTGGACACTGTGTGAGTGAAGTGAAGGACAAAATGTTTC 958  
Db 997 CAGACAGATCTGGCATCTCCACTGGACACTGTGTGAGTGAAGTGAAGGACAAAATGTTTC 1056  
Qy 959 AAGTGGTGGAGTCCCGCTTGTAGACAGCTCCATCTCGTCTCTTACTTACCTTTGG 1018  
Db 1057 AAGTGGTGGAGTCCCGCTTGTAGACAGCTTCTATCCCGCTCTCCATATTTACCTTTGG 1116  
Qy 1019 CTGTACCAAGAACCTTGCGAGATCGACTCTCAGAGTCCATGCTGATCTCTGAGTGTGGT 1078  
Db 1117 CTGTACCAAGAACCTTGCGAGATCGACTTGTAGAGTGTGATGCTGACCTCGAGTGTGGT 1176  
Qy 1079 GGGTATCCAGTTTGTCAAAATCTTGTATCCGTCACAACTTTGGCTGGAAAGGAAATAG 1138  
Db 1177 GGGTGTCTCAGTTTGTCAAAATCTTGTATCCGTCACAGCTTGGCTAGAAAAGAAATAG 1236  
Qy 1139 AAGAAACCAACCAAGAGCTTGGCTTCAAAATCTCAGTATTTGGAGTCCATGTGACAGCA 1198  
Db 1237 AAGAAAGCCACCAAGAGCTTGGCTTCAAAATCTCAGTATTTGGAGTCCATGTGACAGCA 1296  
Qy 1199 CTGCAAAAGTGGGACAGAGCAGCTTCCATCCATTTGAGGAATACATGTTGACAGTGTG 1258  
Db 1297 CAGCAAAAGTGGGACAGAGCAGCTTCCATCCATTTGAGGAATACATGTTGAGTGTG 1356  
Qy 1259 AAGAAACATTTTCAGCTTCTCGAACGAGAAATGAAAGTGGATTAAGAAAGAGTGTATCTGG 1318  
Db 1357 AAGAAACATTTTCAGCTTCTCGAACGAGAAATGAAAGTGGATTAAGAAAGAGTGTATTTGG 1416

Qy 1319 CCAGTATGACCCCTTCTTTGTTAAAGGAGGCAAGACAAAGTACTCCAATTTATGAATTTA 1378  
Db 1417 CCAGATGATCCCTTCTTTTAAAGGAGGCAAAACAAAGTACCCCAATTTATGAATTTA 1476  
Qy 1379 TTAGTGAATACTCTATTTCTTTGTTGTCAGTGGATACACACCGGATACACAGAAATTCAC 1438  
Db 1477 TTAGTGAATACTCTATTTCTTTGTTGTCAGTGGATGCAATTCGATACACAGAAATTCAC 1536  
Qy 1439 TTGGGGGCGTATCTCTGATATACACTTCTCTCCAGGCTGACTTCTTGTGTGTACTT 1498  
Db 1537 TTCTGGAGTATCTCTGGATATACATTTCTCTCAGGCGAGACTTCTTAGTGTGTACTT 1596  
Qy 1499 TTTTCATCCAGGCTCTGAGGCTTCTTATGAAATCATGCAAAACATGCTCATCTCTGATGCT 1558  
Db 1597 TTTTCATCCAGGCTCTGAGGCTTCTTATGAAATCATGCAAAACATGCTCATCTCTGATGCT 1656  
Qy 1559 CTGCAAACTTCCATCTTTTAGATGATCATCTACTATTTTGGAGGCAAAATGCCCAACACC 1618  
Db 1657 CTGCAAACTTCCATCTTTTAGATGATCATCTACTATTTTGGGGCCAGAAATGCCCAACATC 1716  
Qy 1619 AGATTGAGTGTATCTCTCAACCACTCGAACTAAAGAGGAAATCCCATGGAACTTGAG 1678  
Db 1717 AAATTGGCATTTATGTTCAACCACTCGAACTGAGATGAAATTTCCCATGGAACTTGAG 1776  
Qy 1679 ATATCATTTGGTGTGGTGGAAACCATTTGGAATGTTACTCTTAAAGGTGTCAACAGAAAC 1738  
Db 1777 ATATCATTTGGTGTGGTGGAAATCATTTGGGATGCTATTTAAAGGTGTCAACAGGAAAT 1836  
Qy 1739 TAGGAAACAGGCGCTGTACCCCTTCTACAAAGTCCGAGAGAGATAGAAACAGTCAAAT 1798  
Db 1837 TGGGAAGGACGGGCTATATCTCCCTCTACAAAGTTCGAGAGAGATAGAAACGCTCAAGT 1896  
Qy 1799 ACCCTACATATCTTGAAGCTGAAATAGAGATGGAGTGTAAAGATTTAAACAGAAAT 1858  
Db 1897 ACCCTACATATCTTGAAGCTGAGAAATTAAGCTCAGATGGAAAGAGATAACGACAAACT 1956  
Qy 1859 TAGTTTCAGACCATCTCAG 1876  
Db 1957 CAGTTTCAGCAACCACTCAG 1974

## RESULT 7

ABA98809  
ID ABA98809 standard; cdna; 1759 BP.

XX AC ABA98809;

XX XX 07-MAY-2002 (first entry)

XX DE Alpha1,6-fucosyl transferase cdna.

XX XX Plant; glycoprotein; alpha1,6-fucosyl transferase; alpha1,6-FT;  
KW enzyme; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
CDS 17..1744  
/\*tag= a  
/product= "Alpha1,6-fucosyl transferase"

XX JP2001333787-A.

XX XX 04-DEC-2001.

XX XX 06-MAR-2001; 2001JP-0062704.

XX XX 22-MAR-2000; 2000JP-0081059.

XX PA (TANI/) TANIGUCHI N.

XX PA (SEKI/) SEKI T.

XX PA (FUJI/) FUJIYAMA K.



XX 05-MAR-1998 (first entry)  
DT Pig alpha 1-6 fucosyltransferase gene.  
XX  
DE Alpha 1-6 fucosyltransferase; enzyme; pig; human; fucose transfer;  
KW guanosine diphosphate; sugar chain synthesis; modification; antibody;  
KW GlcNAc; cancer diagnosis; ss.  
XX  
OS Sus scrofa.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..1728  
FT /tag= a  
XX  
XX WO9727303-A1.  
XX  
PD 31-JUL-1997.  
XX  
XX 23-JAN-1997; 97WO-JP00171.  
XX  
PR 22-JUL-1996; 96JP-0192260.  
PR 24-JAN-1996; 96JP-0010365.  
PR 21-JUN-1996; 96JP-0161648.  
PR 24-JUN-1996; 96JP-0162813.  
XX  
XX (TOYM) TOYO BOSEKI KK.  
XX  
XX Shiba T, Taniguchi N, Uozumi N, Yanagidani S;  
PI P-PSDB; AAW22124.  
XX  
XX Human or pig alpha 1-6 fucosyltransferase and DNA encoding it - for  
PT synthesis and modification of sugar chains and used as an antigen  
PT for production of diagnostic antibodies  
XX  
PS Claim 5; Page 30-34; 61pp; Japanese.  
XX  
CC AAT76573 and AAT76574 represent the coding sequences for the pig and  
CC human alpha 1-6 fucosyltransferases of the invention, respectively. The  
CC encoded enzyme transfers fucose from guanosine diphosphate to the  
CC 6-hydroxyl group of the GlcNAc nearest to R in the receptor molecule:  
CC (GlcNAc)beta 1-2Man(alpha 1-6)(GlcNAc)beta 1-2Man(alpha 1-3)Man(beta  
CC 1-4GlcNAc)beta 1-4GlcNAc-R to give (GlcNAc)beta 1-2Man(alpha 1-6)(GlcNAc)beta  
CC 1-2Man(alpha 1-3)Man(beta 1-4GlcNAc)beta 1-4(Fu)alpha 1-6(GlcNAc)-R. It has  
CC an optimum pH of about 7.0 (pig) or 7.5 (human), and is stable over the  
CC pH range 4-10 after 5 hours at 4 degrees C. The optimum working  
CC temperature of the enzyme is 30-37 degrees C. A bivalent metal is not  
CC required for activity of the enzyme, and the enzyme is not inhibited in  
CC the presence of 5 mM EDTA. The enzyme is useful in the synthesis and  
CC modification of sugar chains, and as antigen for the production of  
CC antibodies recognising the enzyme. The antibodies can be used for the  
CC diagnosis of cancer and other diseases.  
XX  
SQ Sequence 1728 BP; 521 A; 362 C; 419 G; 426 T; 0 other;  
  
Query Match 73.4%; Score 1474.2; DB 18; Length 1728;  
Best Local Similarity 90.9%; Pred. No. 0;  
Matches 1569; Conservative 0; Mismatches 158; Indels 0; Gaps 0;  
  
QY 100 ATGCGGGCATGGACGTGTTCCCTGGGTTGGATTATGCTCATCTTTTGGCTGGGGACC 159  
DB 1 ATGCGGCCCATGGACGTGTTCCGCGTTGGATTATGCTCATCTTTTGGCTGGGGACC 60  
  
QY 160 TTATTTGTTTATATAGTGGTCAATTGGTTCCAGATAATGACCACCTTGACCATCTAGC 219  
DB 61 TTGCTATTTTACATAGTGGTCACTTGGTACAGATAATGACCACCTTGATCACTCTAGC 120  
  
QY 220 AGAGAACTCTCAAGATTCTTCCAAAGCTGGAGCGCTTAAACAAACAAATGAAGACTTG 279  
DB 121 CGAGAACTGCTCCAAGATTCTTGGCAAAGCTGGAAAGCTTAAACAAACAAATGAAGACTTG 180

QY 280 AGGAGAAATGGCTAGTCTCTCCGAATACCGAAGCGCCCTATTGTATCAGGGGACAGCTACA 339  
DB 181 AGGAGAAATGGCTGAATCTCTCCGAATACCGAAGCGCCCTATTGTATCAGGGGACAGCTTCA 240  
  
QY 340 GGAAGAGTCCGTGTTTATGAAGAACAGCTTGTAAAGGCCAAAGAACAGATTTGAAATTTAC 399  
DB 241 GGAAGAGTCCGTGTTTATGAAGAACAGATTTATGAAGGCCAAAGAACAGATTTGAAATTTAT 300  
  
QY 400 AAGAAACAGCTAGGAATGATCTGGAAAGAGATCATGAAATCTTTAAGGAGGAGATTGAA 459  
DB 301 AAGAAACAACTAATAATGGTCCAGGAGAGATCATGAAATCTTAAAGGAGGAGATTGAA 360  
  
QY 460 AATGAGCTAAAGAGCTCTGTTTTTTCTACAAAGTGAATTTGAAGAAATTTAAAGAAATTA 519  
DB 361 AATGAGCTAAAGAGCTCTGTTTTTTCTACAAAGTGAATTTGAAGAAATTTAAAGAAATTTA 420  
  
QY 520 GAAGGAAACGAATCTCCAAAGACATGCAGATGAAATTTCTTTTGGATTATAGGACATCATGAA 579  
DB 421 GAAGGAAATGAATCTCCAAAGACATGCAGATGAAATTTCTATCAGATTTGGGACATCATGAA 480  
  
QY 580 AGGTCTATCATGACAGATCTATCTACTCTCAGTCCAAACAGATGGAGCAGGTGAGTGGCGG 639  
DB 481 AGGTCTATATGACGATCTATCTACTCTCAGTCCAAACAGATGGGCGAGGTGATTTGGCGT 540  
  
QY 640 GAAAGAAAGCCAAAGATCTGACAGAGCTGCTCCAGCGGAGAAATAACATATCTTCAGAAAT 699  
DB 541 GAAAGGAGGCCAAAGATCTGACAGAGCTGCTCCAGCGGAGAAATAACATATCTTCAGAAAT 600  
  
QY 700 CCCAAGGACTGCAGTAAAGCCAGAAAGCTGTATGTATATATCAACAAAGGCTGTGGCTAT 759  
DB 601 CCCAAGGACTGCAGTAAAGCCAGAAAGCTGTATGTATATCAACAAAGGCTGTGGCTAT 660  
  
QY 760 GGATGTCAACTCCATCATGTGTTTACTGCTTCTATGATTTGCTTATGGCACCAGCGGAACA 819  
DB 661 GGCTGTGAGCTCCATCATGTAGTGTACTGCTTTATGATTTGATATGGCACCAGCGGAACA 720  
  
QY 820 CTCACTTTGGAATCTCAGAAATGGCGCTATGCTACTTGGAGGATGGGAGCTGTGTTTAGA 879  
DB 721 CTGCGCTTGGAACTCTCACAATTTGGCGCTACGCTACTGGGGATGGGAACTGTGTTTAGA 780  
  
QY 880 CCTGTAGTGAAGACATGACAGACAGAGCTGTGGCTCTCCACTGGACACTGTGTAGGTGAA 939  
DB 781 CCTGTAGTGAAGACATGACAGACAGAGCTGTGGCGAGCTCCACTGGACACTGTGTAGGTGAA 840  
  
QY 940 GTGAGGACAAAATGTTCAAGTGTGAGCTGCTCCCAATTTGATAGACAGCTCCATCTCGT 999  
DB 841 GTAAAGGACAAAATGTTCAAGTGTGAGCTGCTCCCAATTTGATAGACAGCTGTGTATCTCGT 900  
  
QY 1000 CCTCTTACTTTACCCCTTGGCTGTATCCAGAAAGACCTTGCAGATCGACTCCTGAGAGTCCAT 1059  
DB 901 CCTCTTACTTTACCCCTTGGCTGTATCCAGAAAGACCTTGCAGATCGACTGTGTAGAGTCCAT 960  
  
QY 1060 GGTGATCTCGAGTGTGGTGGTATCCAGTTTGTCAAATATCTGTATCCGTCCCAACCT 1119  
DB 961 GGTGATCTCGAGTGTGGTGGTATCCAGTTTGTCAAATATCTGTATCCGTCCCAACCT 1020  
  
QY 1120 TGGCTGGAAGGGAATAGAAAGAACCCACCAAGAGCTTGGCTTCAACATCCAGTTAT 1179  
DB 1021 TGGCTGGAAGGGAATAGAAAGAGGCCACCAAGAGCTTGGCTTCAACATCCAGTTAT 1080  
  
QY 1180 GGAGTCCATGTCAGACGCTGACAAAGTGGGAAACAGAGCAGCTTCCATCCCATTCAG 1239  
DB 1081 GGAGTCCATGTCAGACGCTGACAAAGTGGGAAACAGAGCAGCTTCCATCCCATTCAG 1140  
  
QY 1240 GAATACATGTTACGTTGAAGAAACAATTTTCAGCTTCTCGAAGCAGAAATGAAAGTGGAT 1299  
DB 1141 GAATACATGTTACGTTGAAGAAACAATTTTCAGCTTCTCGAAGTTCGAGAAATGCAAGTGGAT 1200  
  
QY 1300 AAAAAAGAGTGTATCTGGCCACTGTAGCCCTTCTTTTGTAAAGGAGGAGGAGCAAG 1359  
DB 1201 AAAAAAGAGTGTATCTTTGGCCACAGATGACCTGCTTTGTTTAAAAAGAGGAGGAGCAAG 1260  
  
QY 1360 TACTCCATTTATGAATTTATTTAGTGATAACTCTATTTCTTGGTCAGCTGGACTACACAAC 1419

1261	Db		TACCCAGTTATGAATTTATTAGTGTAACCTATCTCTTTGGTCAGCTGACATACATAAT	1320
1420	Qy		CGATACACAGAAATTCACITTCGGGGCGTCATCTCGATATACACITTTCTCTCCAGGCT	1479
1321	Db		CGATATACAGAAATTCACITTCGGGGGTGATCTCGATATACACITTTCTCTCCAGGCA	1380
1480	Qy		GACTTCCTCTGTGTACTTTTTTCATCCAGGCTCTGTAGGGTGTGCTTATGAAATCATGCAA	1539
1381	Db		GACTTCTAGTGTACTTTTTTCATCGAGTCTGTAGAGTTCGTTATGAAATCATGCAA	1440
1540	Qy		ACACTGCATCCTGATGCCTCTGCAAACTTCCATTCCTTTAGATGACATCTACTATTTTGG	1599
1441	Db		GCCTGCATCCTGATGCCTCTCGCACTTCGGTCTTTTGGATGACATCTACTATTTTGG	1500
1600	Qy		GGCCAAAATGCCCAACACAGATGCAAGTTTATCTCTCAACCACTCGAACTAAAGAGAA	1659
1501	Db		GGCCCAAATGCCCAACCAAAATGGCCATTTATCTCACCAACCTCGAACTGAAGAGAA	1560
1660	Qy		ATCCCCATGAACTCGAGATATCATTTGGTGTGCTGGAAACCATTTGGAATGCTACTCT	1719
1561	Db		ATCCCCATGAACTCGAGATATTTATTTGGTGTGCTGGAAATCACTGGGATGCTATCCT	1620
1720	Qy		AAAGGTGTCACACAGAAAACTTAGGAAAAACAGGCGCTGTACCCCTTCTACAAAGTCCGAG	1779
1621	Db		AAAGGTGTTAACAGAAAACTGGGAGGACGGGCTATATCCCTCCTACAAAGTTCGAG	1680
1780	Qy		AGATGAGAAACAGTCAAAATACCTTACATATCTCTGGAAGCTGAAAAATA	1826
1681	Db		AGATGAGAAACAGTCAAGTACCCCAATATCCCGAGCTGACAAGTA	1727

RESULT 9	
AAC63891	
ID AAC63891	standard; cDNA; 979 BP.
XX	
XX AAC63891;	
XX	
XX AC	
XX DT	
XX 09-FEB-2001	(first entry)
XX	
XX Chinese hamster	FUT8 cDNA.
XX	
XX Immunologically functional molecule;	immune system; immunomodulation;
KW glycosylation; fucose;	N-acetylglucosamine; cancer; circulatory disease;
KW viral infection; bacterial infection;	allergy; autoimmune disease;
KW inflammation; antibody; Chinese hamster;	FUT8; ss.
XX	
XX Cricetulus griseus.	
OS	
XX	
XX WO200061739-A1.	
PN	
XX	
XX 19-OCT-2000.	
PD	
XX	
XX 07-APR-2000;	2000WO-JP02260.
PP	
XX	
XX 09-APR-1999;	99JP-0103158.
PR	
XX	
XX (KYOW ) KYOWA HAKKO KOGYO KK.	
PA	
XX	
XX Hanai N, Nakamura K, Shoji E, Yamasaki M, Uchida K, Shinkawa T;	
PI Imabeppu S, Kanda Y, Yamane N, Anazawa H;	
PI	
XX	
XX WPI; 2000-665129/64.	
DR	
XX	
XX Control of activity of antibodies and other immunologically functional	
PT molecules by addition or removal of fucose from sugar chain for	
PT diagnosis and treatment of cancer, allergy and other diseases -	
PT	
XX	
XX Example 8; Page 72-73; 81pp; Japanese.	
PS	
XX	
XX The invention relates to a method for controlling the activity of an	
CC immunologically functional molecule (e.g., an antibody) where the	
CC control is effected by the presence or absence of fucose bound to an	
CC	



```

XX 09-FEB-2001 (first entry)
XX Rat FUT8 cDNA.
XX Immunologically functional molecule; immune system; immunomodulation;
KW glycosylation; fucose; N-acetylglucosamine; cancer; circulatory disease;
KW viral infection; bacterial infection; allergy; autoimmune disease;
KW inflammation; antibody; rat; FUT8; ss.
XX Rattus sp.
XX WO200061739-A1.
XX 19-OCT-2000.
XX 07-APR-2000; 2000WO-JP02260.
XX 09-APR-1999; 99JP-0103158.
XX (KYOW ) KYOWA HAKKO KOGYO KK.
XX Hanai N, Nakamura K, Shoji E, Yamasaki M, Uchida K, Shinkawa T;
PI Imabeppu S, Kanda Y, Yamane N, Anazawa H;
XX WPI; 2000-665129/64.
XX Control of activity of antibodies and other immunologically functional
PT molecules by addition or removal of fucose from sugar chain for
PT diagnosis and treatment of cancer, allergy and other diseases -
XX Example 8; Page 73-74; 81pp; Japanese.
XX The invention relates to a method for controlling the activity of an
CC immunologically functional molecule (e.g., an antibody) where the
CC control is effected by the presence or absence of fucose bound to an
CC N-acetylglucosamine residue at the reducing end of the sugar chain on
CC the immunologically functional molecule. The invention also relates to
CC methods for the diagnosis, prevention or treatment of diseases which
CC involve the modified immunologically functional molecule, and agents
CC which stimulate the activity of an immunologically functional molecule.
CC The methods of the invention are used for the diagnosis, treatment and
CC prevention of a broad range of diseases including cancer, circulatory
CC disease, viral or bacterial infection, allergy, autoimmune disease and
CC inflammation. The present sequence represents a rat FUT8 cDNA isolated
CC in an exemplification of the invention.
XX Sequence 979 BP; 283 A; 223 C; 220 G; 253 T; 0 other;
SQ
Query Match 44.5%; Score 893.2; DB 21; Length 979;
Best Local Similarity 94.6%; Pred. No. 4.5e-255;
Matches 925; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY 819 ACTCATCTTGGAAATCTCAGAAATGGCGCTATGCTACTGAGGATGGGAGACTGTTTAG 878
DB 1 ACTCATCTTGGAAATCTCAGAAATGGCGCTATGCTACTGAGGATGGGAGACTGTTTAG 60
QY 879 ACCTGTAAGTGAGATGCACAGACAGGTCTGGCCTCTCCACTGGACACTGGTCAGGTGA 938
DB 61 ACCTGTAAGTGAGATGCACAGACAGATCTGGCCTCTCCACTGGACACTGGTCAGGTGA 120
QY 939 AGTGAAGGACAAATAATGTTTCAAGTGGTGGAGCTCCCACTTGTAGACAGCCTCCATCTCG 998
DB 121 AGTGAATGACAAATAATTCAGTGGTGGAGCTCCCACTTGTAGACAGCCTTCATCTCG 180
QY 999 TCCTCCTTACTTACCCTTGGCTGTACACAGAACCTTGCAGATCGACTCTCGAGGTCCA 1058
DB 181 GCCTCCTTACTTACCCTTGGCTGTTCACAGAACCTTGCAGATCGACTCGTAAGAGTCCA 240
QY 1059 TGGTGATCTCGCAGTGGTGGGTATCCAGTTTGTCAAAATCTTGTATCCCGTCCACAACC 1118
DB 241 TGGTGATCTCGCAGTGGTGGGTATCCAGTTTGTCAAAATCTTGTATCCCGTCCACAACC 300

```

```

QY 1119 TTGGCTGGAAGGGAATAGAGAAACCAACCAAGAGCTTGGCTTCAACATCCAGTTAT 1178
DB 301 TTGGCTGGAAGGGAATAGAGAAAGCCACCAAGAGCTTGGCTTCAACATCCAGTTAT 360
QY 1179 TGGAGTCCATGTCAGAGCAGTCAAAAGTGGGAACAGAACGAGCCTTCCATCCATTGA 1238
DB 361 TGGAGTCCATGTCAGAGCAGCAGCAAAAGTGGGAACAGAGCAGCCTTCCATCCATTGA 420
QY 1239 GGAATACATGTCACACGTTGAAGAACATTTTTCAGCTTCTCGAACGCAAGTGAAGTGA 1298
DB 421 AGAGTACATGTCATGTTGAAGAACATTTTTCAGCTTCTCGAACGCAAGTGAAGTGA 480
QY 1299 TAAAAAAGAGTGTATCTGGCCACTGATGACCCCTTCTTTGTTAAAGGAGGCAAGACAAA 1358
DB 481 TAAAAAAGAGTGTATCTGGCTACCGATGACCCCTTCTTTGTTAAAGGAGGCAAGACAAA 540
QY 1359 GTACTCCAAATTTATGAATTTATTTAGTGATACTCTATTCTTGGTGCAGCTGACATACAAA 1418
DB 541 GTACTCCAAATTTATGAATTTATTTAGTGATACTCTATTCTTGGTGCAGCTGACATACAAA 600
QY 1419 CCGATACACAGAAATTCACCTTCGGGGCGTGATCTCGATATACACTTTCTCTCCAGGC 1478
DB 601 TCGGTACACAGAAATTCACCTTCGGGGCGTGATCTCGATATACACTTTCTCTCCAGGC 660
QY 1479 TGACTTCTTGTGTGTACTTTTTCATCCAGGTCTGTAGGTTGCTTATGAATCATGCA 1538
DB 661 TGACTTCTTGTGTGTACTTTTTCATCCAGGTCTGTAGGTTGCTTATGAATCATGCA 720
QY 1539 AACACTCATCTGATGCTCTGCAAACTTCCATTTCTTTAGATGACATCTACTATTTTGG 1598
DB 721 AACCTGATCTCTGATGCTCTGCAAACTTCCATTTCTTTAGATGACATCTACTATTTTGG 780
QY 1599 AGGCCAAATGCCCAACACAGATTGCGATTATCTCTCACCAACTCGAATCAAGAGGA 1658
DB 781 AGGCCAAATGCCCAACACAGATTGCGATTATCTCTCACCAACTCGAATCAAGAGGA 840
QY 1659 AATCCCATGGAACCTGGAGATATCATTTGTGTGGCTGGAAACCATTTGGAATGTTACTC 1718
DB 841 AATTCCAATGGAACCTGGAGATATCATTTGTGTGGCTGGAAACCATTTGGAATGTTACTC 900
QY 1719 TAAAGGTGTCAACAGAAACCTAGGAAACAGGCTGTACCCCTTCTCAAAAGTCCGAGA 1778
DB 901 TAAAGGTGTCAACAGAAACCTAGGAAACAGGCTGTATATCCCTCTCAAAAGTCCGAGA 960
QY 1779 GAAGATAGAAACAGTCAA 1796
DB 961 GAAGATAGAAACAGTCAA 978

```

RESULT 12  
ABK70037  
ID ABK70037 standard; cDNA; 979 BP.

AC ABK70037;

DT 15-JUL-2002 (first entry)

XX Antibody production method related cDNA #5.

XX Antibody production; cytostatic; immunomodulator; vasotropic; virucide;  
KW antibacterial; antinflammatory; antiallergic; allergy; inflammation;  
KW autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO;  
KW tumour; circulatory disease; infection; primer; ss.

OS Rattus norvegicus.

XX WO200231140-A1.

PN 18-APR-2002.

XX 05-OCT-2001; 2001WO-JP08804.

XX 06-OCT-2000; 2000JP-0308526.



XX (KYOM ) KYOWA HAKKO KOGYO KK.  
XX Kanda Y, Sato M, Nakamura K, Uchida K, Shinkawa T, Yamane N;  
XX Hosaka E, Yamano K, Yamasaki M, Hanai N;  
XX WPI; 2002-340182/37.  
XX Cells producing antibody compositions including antibody fragments and  
XX fusion proteins with Fc domain of antibody, useful for prevention or  
XX treatment of cancer, immune diseases, circulatory diseases and  
XX infections  
XX Example 9; Page 13-14; 314pp; Japanese.  
XX This invention relates to novel method for antibody production  
XX comprising a Chinese hamster ovarian tissue-originated (CHO) cell  
XX transferred with a gene encoding an antibody molecule for producing a  
XX composition comprising an antibody molecule with an Fc domain bonded  
XX to the N-glycoside linkage complex sugar chain. The produced antibody  
XX compositions are drugs for prevention or treatment of diseases  
XX accompanying tumour, allergy or inflammation, autoimmune diseases,  
XX circulatory diseases, and viral and bacterial infections. The  
XX antibodies can be stably produced using the method of the invention  
XX with high binding activity and potency thus leading to high safety and  
XX reduced side effects when applied alone or in combination with other  
XX drugs for therapy. The present sequence represents a nucleotide  
XX molecule used in the method of the invention.  
XX SQ Sequence 979 BP; 283 A; 223 C; 220 G; 253 T; 0 other;  
Query Match 44.5%; Score 893.2; DB 24; Length 979;  
Best Local Similarity 94.6%; Pred. No. 4.5e-255;  
Matches 925; Conservative 0; Mismatches 53; Indels 0; Gaps 0;  
QY 819 ACTCATCTTGAATCTCAGAAATGGCGTATGCTACTGAGGATGGGAGACTGTTTGTAG 878  
DB 1 ACTCATCTTGAATCTCAGAAATGGCGTATGCTACTGAGGATGGGAGACTGTTTGTAG 60  
QY 879 ACCTGTAAGTGAGACATGACAGACAGAGTGGCGCTCTCCACTGAGACACTGGTCAGGTGA 938  
DB 61 ACCTGTAAGTGAGACATGACAGACAGAGTGGCGCTCTCCACTGAGACACTGGTCAGGTGA 120  
QY 939 AGTGAAGACAAATATTTCAAGTGGTGGAGTGGCGCTCTCCACTGAGACACTGGTCAGGTGA 998  
DB 121 AGTGAAGACAAATATTTCAAGTGGTGGAGTGGCGCTCTCCACTGAGACACTGGTCAGGTGA 180  
QY 999 TCCTCTTACTTACCTTGGCTGTACAGAGACCTTGCAGATCGACTCTCGAGAGTCCA 1058  
DB 181 GCCTCTTACTTACCACTGGCTGTTCAGAGACCTTGCAGATCGACTCGTAAGAGTCCA 240  
QY 1059 TGGTGATCTCGAGTGGTGGGTATCCAGTTTGTCAAATTTGATTCGCTCCCAACACC 1118  
DB 241 TGGTGATCTCGAGTGGTGGGTATCCAGTTTGTCAAATTTGATTCGCTCCCAACACC 300  
QY 1119 TTGGCTGGAAGGGAATAGAGAAACCAAGAGAGCTTGGCTTCAACATCCAGTTAT 1178  
DB 301 TTGGCTGGAAGGGAATAGAGAAACCAAGAGAGCTTGGCTTCAACATCCAGTTAT 360  
QY 1179 TGGAGTCCATGTGACAGCGCTGACAAAGTGGGAACAGAGCGCTTCCATCCCATTTGA 1238  
DB 361 TGGAGTCCATGTGACAGCGCTGACAAAGTGGGAACAGAGCGCTTCCATCCCATTTGA 420  
QY 1239 GGAATACATGTTACAGTTGGAAGACATTTTCAGTTTGTCCGAAACGAGAAAGTGGGA 1298  
DB 421 AGAGTACATGTTACAGTTGGAAGACATTTTCAGTTTGTCCGAAACGAGAAAGTGGGA 480  
QY 1299 TAAAGGAGAGTGTATCTGGCCACTGATGACCCCTTCTTTGTTAAAGGAGGCAAGACAAA 1358  
DB 481 TAAAGGAGAGTGTATCTGGCCACTGATGACCCCTTCTTTGTTAAAGGAGGCAAGACAAA 540  
QY 1359 GTACTCCCAATTATGATTTATTAGTGATACTCTATTTCTTGGTCAGCTGGACTACACAA 1418

DB 541 GTACTCCCAATTATGATTTATTAGTGATACTCTATTTCTTGGTCAGCTGGACTACACAA 600  
QY 1419 CCGATACACAGAAATTCACCTTCGGGGCGTGATCCTCGATATACACTTTCTCTCCAGGC 1478  
DB 601 TCGGTACACAGAAATTCACCTTCGGGGCGTGATCCTCGATATACACTTTCTCTCAGGC 660  
QY 1479 TGACTTCTCTTGTGTGTACTTTTTCATCCAGGCTGTAGGGTGTCTTATGAAATCATGCA 1538  
DB 661 TGACTTCTCTTGTGTGTACTTTTTCATCCAGGCTGTAGGGTGTCTTATGAAATCATGCA 720  
QY 1539 AACACTGATCTGATGCTCTCGAAACTTCCATCTTTAGATGACATCTACTATTTTGG 1598  
DB 721 AACCTGCTGCTGATGCTCTCGAAACTTCCATCTTTAGATGACATCTACTATTTTGG 780  
QY 1599 AGGCAAAATGCCCAACACAGATTGCGATTATCTCCACCACTCGAACTAAAGAGA 1658  
DB 781 AGGCAAAATGCCCAACACAGATTGCGATTATCTCCCACTCGAACTGATGAGA 840  
QY 1659 AATCCCATGGAACCTCGAGATATCATTTGGTGGCTGGAAACCATTTGGATGTTACTC 1718  
DB 841 AATCCCATGGAACCTCGAGATATCATTTGGTGGCTGGAAACCATTTGGATGTTACTC 900  
QY 1719 TAAAGGTGTCACAGAAACTAGGAAACAGGCTGTACCTTCTCTACAAAGTCCGAGA 1778  
DB 901 TAAAGGTGTCACAGAAACTAGGAAACAGGCTGTATCCCTCTCTACAAAGTCCGAGA 960  
QY 1779 GAAGATAGAAACAGTCAA 1796  
DB 961 GAAGATAGAAACAGTCAA 978  
RESULT 13  
AAF87952  
ID AAF87952 standard; DNA; 1017 BP.  
XX AAF87952;  
XX AC AC  
XX 20-JUL-2001 (first entry)  
XX Human alpha 1-6 fucosyltransferase 237-575 encoding DNA SEQ ID NO:1.  
DE Human; alpha 1-6 fucosyltransferase; alpha 1-6 fucT; antibody;  
XX alpha 1-6 fucose transferase; anti-human alpha 1-6 fucT; immunoassay;  
KW ds.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
FT CDS 1..1017  
FT /\*tag= a  
FT /partial  
FT /product= "alpha 1-6 fucosyltransferase 237-575"  
XX JP2001011097-A.  
XX 16-JAN-2001.  
XX 29-JUN-1999; 99JP-0183569.  
XX 29-JUN-1999; 99JP-0183569.  
XX (FURE ) FUJIREBIO KK.  
XX WPI; 2001-275926/29.  
XX P-PSDB; AAB75061.  
XX Novel anti-human alpha1-6 fucose transferase antibody useful for  
XX immunoassay -  
XX Example 1; Page 6-7; lipp; Japanese.  
XX The present invention describes an anti-human alpha 1-6  
XX fucosyltransferase (alpha 1-6 fucT, also called alpha 1-6 fucose

CC transferase) antibody. Also described are: (1) a hybridoma producing  
CC the above monoclonal antibody; (2) an immunoassay for detecting human  
CC alpha 1-6 fucT by using the above antibody or its antibody fragment;  
CC and (3) a reagent used for the above immunoassay. The anti-human  
CC alpha 1-6 fucT antibody can be used for immunoassay. The present  
CC sequence encodes the human alpha 1-6 fucT protein of residues 237 to  
CC 575, which is used in an example from the present invention.  
XX  
SQ Sequence 1017 BP; 296 A; 231 C; 228 G; 262 T; 0 other;

Query Match 43.9%; Score 881; DB 22; Length 1017;  
Best Local Similarity 91.6%; Pred. No. 28-251; Mismatches 85; Indels 0; Gaps 0;  
Matches 932; Conservative 0;

QY 808 ACCACGCGAACACTCATCTTGGAACTCAGAAATGGCGCTACTGCTACTGGAGATGGAG 867  
DB 1 ACCACGCGAACACTCATCTTGGAACTCAGAAATGGCGCTACTGCTACTGGAGATGGAG 60

QY 868 ACTGTTTGTAGACCTGTAAGTGAGACATGCACAGACAGTCTGGGCTCTCCACTGGACAC 927  
DB 61 ACTGTTTGTAGACCTGTAAGTGAGACATGCACAGACAGTCTGGGCTCTCCACTGGACAC 120

QY 928 TGGTCAGGTGAAGTGAAGACAAAATGTTCAAGTGGTGGAGTCTCCACTGTAGACAGT 987  
DB 121 TGGTCAGGTGAAGTGAAGACAAAATGTTCAAGTGGTGGAGTCTCCACTGTAGACAGT 180

QY 988 CTCCATCTCTGCTCTCTCTTACTTACCTTGGCTGTACCAAGACAGCTTGCAGATGACATC 1047  
DB 181 CTTCATCCCGTCTCTCTCTTACTTACCTTGGCTGTACCAAGACAGCTTGCAGATGACATC 240

QY 1048 CTGAGAGTCCATGTTGATCTCTGAGTGGTGGGATCCAGTGTGTCAAATACCTGATC 1107  
DB 241 GTACGAGTGCATGGTGCACCTGCAGTGGTGGGATCTCTGAGTGTGTCAAATACCTGATC 300

QY 1108 CGTCACACCTTGGCTGGAAGGGAATAGAGAAACACCAAGAGCTTGGCTTCAA 1167  
DB 301 CGCCACACCTTGGCTGGAAGGGAATAGAGAAACACCAAGAGCTTGGCTTCAA 360

QY 1168 CATCCAGTTATTTGGAGTGCATCTCAGACGCACTGACAAAGTGGGAACAGAGCAGCTTC 1227  
DB 361 CATCCAGTTATTTGGAGTGCATCTCAGACGCACTGACAAAGTGGGAACAGAGCAGCTTC 420

QY 1228 CATCCAGTTAGGAATACATGTTGACATGTTGAAGAACATTTTCAGCTTCTCGAAGCAGA 1287  
DB 421 CATCCAGTTAGGAATACATGTTGACATGTTGAAGAACATTTTCAGCTTCTCGAAGCAGA 480

QY 1288 ATGAAAGTGAATAAAAGAGTGTATCTGGCAGTGTATTTGGCCACAGATGCCCTTCTTTAAGGAG 1347  
DB 481 ATGAAAGTGAATAAAAGAGTGTATTTGGCCACAGATGCCCTTCTTTAAGGAG 540

QY 1348 GCAAAAGACAAAGTACTCCAAATATGAATTTATTAGTGATACTCTATTTCTTGGTCAGCT 1407  
DB 541 GCAAAAGACAAAGTACTCCCAATATGAATTTATTAGTGATACTCTATTTCTTGGTCAGCT 600

QY 1408 GGACTACACAAACGATACACAGAAATTCACCTCGGGCGGTGATCCTCGATATACATTT 1467  
DB 601 GGACTACACAAACGATACACAGAAATTCACCTCGGGCGGTGATCCTCGATATACATTT 660

QY 1468 CTCTCCAGGCTGACTTCTTGTGTACTTTTTCATCCAGGCTGTAGGGTGTCTTAT 1527  
DB 661 CTCTCTCAGGCGAGATCTCTAGTGTGTACTTTTTCATCCAGGCTGTAGGGTGTCTTAT 720

QY 1528 GAAATCATGCAAAACACTGCTCTGATGCTCTGCAAACTTCCATTTCTTAGATGACATC 1587  
DB 721 GAAATCATGCAAAACACTGCTCTGATGCTCTGCAAACTTCCATTTCTTAGATGACATC 780

QY 1588 TACTATTTTGGAGGCGCAAAATGCCCAACACAGATTCGAGTTTATTCCTCAACACCTCGA 1647  
DB 781 TACTATTTTGGAGGCGCAAAATGCCCAACACAGATTCGAGTTTATTCCTCAACACCTCGA 840

QY 1648 ACTAAGAGGAAATCCCATGGAACCTCGAGATATCAATTTGGTGTGGCTGGAACCATTTGG 1707  
DB 841 ACTGAGATGAATTTCCCATGGAACCTCGAGATATCAATTTGGTGTGGCTGGAACCATTTGG 900

QY 1708 AATGGTTACTTAAGGTGTCAACAGAAAACACTAGAAAACAGCGCTGTACCTTCTTAC 1767  
DB 901 GATGGCTATTCTTAAGGTGTCAACAGAAAACACTAGAAAACAGCGCTGTATATCTTCTTAC 960

QY 1768 AAAGTCCGAGAGATAGAAAACAGTCAATACCTCATATCTCTGAAGCTGAAAAA 1824  
DB 961 AAAGTCCGAGAGATAGAAAACAGTCAATACCTCATATCTCTGAAGCTGAAAAA 1017

RESULT 14  
AAF87953  
ID AAF87953 standard; DNA; 699 BP.  
XX  
AC AAF87953;  
XX  
DT 20-JUL-2001 (first entry)  
XX  
DE Human alpha 1-6 fucosyltransferase 343-575 encoding DNA SEQ ID NO:2.  
XX  
KW Human; alpha 1-6 fucosyltransferase; alpha 1-6 fucT; antibody;  
KW alpha 1-6 fucose transferase; anti-human alpha 1-6 fucT; immunoassay;  
KW ds.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..699  
FT /\*tag= a  
FT /partial  
FT /product= "alpha 1-6 fucosyltransferase 343-575"  
XX  
PN JP2001011097-A.  
XX  
PD 16-JAN-2001.  
XX  
PF 29-JUN-1999; 99JP-0183569.  
XX  
PR 29-JUN-1999; 99JP-0183569.  
XX  
PA (FURE ) FUJIREBIO KK.  
XX  
DR WPI; 2001-275926/29.  
DR P-PSDB; AAB/5062.  
XX  
PT Novel anti-human alpha1-6 fucose transferase antibody useful for  
XX immunoassay -  
XX  
PS Example 1; Page 7-8; 11pp; Japanese.  
XX  
CC The present invention describes an anti-human alpha 1-6  
CC fucosyltransferase (alpha 1-6 fucT, also called alpha 1-6 fucose  
CC transferase) antibody. Also described are: (1) a hybridoma producing  
CC the above monoclonal antibody; (2) an immunoassay for detecting human  
CC alpha 1-6 fucT by using the above antibody or its antibody fragment;  
CC and (3) a reagent used for the above immunoassay. The anti-human  
CC alpha 1-6 fucT antibody can be used for immunoassay. The present  
CC sequence encodes the human alpha 1-6 fucT protein of residues 343 to  
CC 575, which is used in an example from the present invention.  
XX  
SQ Sequence 699 BP; 220 A; 152 C; 147 G; 180 T; 0 other;

Query Match 29.9%; Score 599.8; DB 22; Length 699;  
Best Local Similarity 91.1%; Pred. No. 8.8e-168;  
Matches 637; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1126 GAAAGGAATAGAGAAACCAACAGAGCTTGGCTTCAACATCCAGTTATTTGGAGTC 1185  
DB 1 GAAAGGAATAGAGAAACCAACAGAGCTTGGCTTCAACATCCAGTTATTTGGAGTC 60

QY 1186 CATGTCCAGACGCTGACAAAGTGGGAACAGAGCAGCTTCCATCCCATTTGAGGAATAC 1245  
DB 61 CATGTCCAGACGCTGACAAAGTGGGAACAGAGCAGCTTCCATCCCATTTGAGGAATAC 120



1246 ATGGTACAGCTTGAAGAACATTTTCAGCTTCTCGAACGACAGATGAAAGTGGATAAAAA 1305  
121 ATGGTGCATGTTGAAGACATTTTCAGCTTCTCGAACGACAGATGAAAGTGGATAAAAA 180  
1306 AGAGTGTATCGGCCACATGATGACCCCTTTTGTAAAGAGGCGGCAAAAGTACTGCC 1365  
181 AGAGTGTATTTGGCCACAGATGACCCCTTTTGTAAAGAGGCGGCAAAAGTACTGCC 240  
1366 AATTATGAATTTATAGTGAATCTATTTCTTGGTCACTGGGACTGACAAACCGATAC 1425  
241 AATTATGAATTTATAGTGAATCTATTTCTTGGTCACTGGGACTGACAAACCGATAC 300  
1426 ACAGAAATTCACCTTCGGGGCGTATCTCGATATACACTTTCTCTCCAGGCTGACTTC 1485  
301 ACAGAAATTCACCTTCGGAGTATCTCGATATACACTTTCTCTCCAGGCTGACTTC 360  
1486 CTTGCTGATCTTTTCAATCCAGCTCTGTAGGGTGTCTATGAAATCATGCAACACTG 1545  
361 CTAGTGTGATCTTTTCAATCCAGGCTCTGTAGGGTGTCTATGAAATCATGCAACACTA 420  
1546 CATCTGATGCTCTGCAAACTTCCATTTCTTAGATGACATCTACTATTTGGAGCCAA 1605  
421 CATCTGATGCTCTGCAAACTTCCATTTCTTAGATGACATCTACTATTTGGAGCCAG 480  
1606 AATGCCCAACACAGATTGCAATTTATCTCACCACCTCGAACTTAAAGAGCAATCCCC 1665  
481 AATGCCCAACATCAATTTGCCATTTATGCTCACCAACCCGAACTGCGAGATGAAATCCC 540  
1666 ATGGAACCTGGAGATATCATTTGGTGTGGTGGAAACCAATTTGGAATGTTACTTAAAGGT 1725  
541 ATGGAACCTGGAGATATCATTTGGTGTGGTGGAAATCAATTTGGAATGTTACTTAAAGGT 600  
1726 GTCAACAGAAACTAGGAAACACAGCTGTACCTTCTTCAACAAAGTCCGAGAGAGATA 1785  
601 GTCAACAGAAATTTGGAGAGGACGGGCTTATATCCCTCTCTACAAAGTTCGAGAGAGATA 660  
1786 GAAACAGTCAATACCTTACATATCTTGAAGCTGAAAA 1824  
661 GAAACGTCAGTACCCACATATCTGAGGCTGAGAAA 699

## RESULT 15

ABL04601  
ID ABL04601 standard; cDNA; 2761 BP.

XX AC ABL04601;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 8285.

XX KW Drosophila; developmental biology; cell signalling; insecticide;  
XX KW pharmaceutical; gene; ss.

XX OS Drosophila melanogaster.

XX PN WO2001/1042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEXE ) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR P-PSDB; ABB60498.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions -

PS Claim 1; SEQ ID NO 8285; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
CC sequences (AB101840-AB116175) and the encoded proteins  
CC (AB157737-AB172072).

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 2761 BP; 657 A; 731 C; 755 G; 618 T; 0 other;

Query Match 18.3%; Score 367.6; DB 23; Length 2761;

Best Local Similarity 56.7%; Pred. No. 2.7e-98;

Matches 704; Conservative 0; Mismatches 529; Indels 9; Gaps 1;

QY 572 ATCATGAAGGTCTATCATGACAGATCTATACTACTCTCAGTCAACAGATCGAGCAGTG 631

DB 1083 AGCACAGCGCTCCCTGCTGAGCGATATGGAGCGATGCGTCAGTCCGATGGTACGAGG 1142

QY 632 AGTGGCGGAAAAAGAACAGCAATCTGACAGAGCTGGTCCAGGGGAGAAATACATATC 691

DB 1143 CCTGGCGACACAGAGGACGCGGATTTGAGCGATTTGGTACAGCGCGCTGCACCATC 1202

QY 692 TGCAGATCCCAAGSACTGACAGCAAGCCAGAAAGTGGTATGTAATCAACAAAGGCT 751

DB 1203 TACAGATCCCAAGTATGTCAGAAATGCTCCAGAGTGGTCTGTAAGCTCAACAGGCT 1262

QY 752 GTGGCTATGGAATGTCACCTCCATCATGTGGTCTTACTGCTTCATGATGCTTTATGSCACC 811

DB 1263 GTGGCTACGGCTGCCAACTGCACCATGTGGTGTATGCTTTATAGTGGCTACGCCACCG 1322

QY 812 AGCGAACACTCATCTTGGAAATCTCAGAAATTTGGCGCTATGCTACTGGAGGATGGGAGACTG 871

DB 1323 AGCGAGCGCTCATCTTAAATCCCGGATGGCGGTATCACAGGGCGGCTGGGAGGAGG 1382

QY 872 TGTTTACACCTGTAAAGTGAGACATGCACAGACAGTCTGGGCTCTTCCACTGGACACTGGT 931

DB 1383 TGTTCCAAACCGGTGTCCAAACAGCTGCCATGATGCGGCGACCCCAACACATACAAATTGCG 1442

QY 932 CAGGTGAAGTGAAGGACAAATAATGTTCAAGTGGTTCGAGCTCCCATTTGTAGACAGCTCC 991

DB 1443 CGGGCAAGCCGAACACAC-----CAGGTGCTGGTGTGTCCTCATCATTCGCTGA 1493

QY 992 ATCTGTGCTCTCTTACTTACCTTGGTGTGACAGAAAGACCTTGCAGATCGACTCTCTGA 1051

DB 1494 TCCGAGACCCCGCTTACCTGCGCTGCGCGTTCGCGAGGATCTGGCGCGGAGACTTAAGC 1553

QY 1052 GAGTCCATGGTGTATCCCTGACGTGGTGGGTATCCAGTGGTGTCAAAATCTTATGTCGCTC 1111

DB 1554 GTCTGATGGAGAGACCCATCTGTTGGTGGGTGGGTGAGTCTTCCCTTAAGTATTTGCTCCGAC 1613

QY 1112 CACAACTTGGCTGGAAAGGGAATAGAAAGAAACACCAAGAGCTTGGCTTCAAAACATC 1171

DB 1614 CGCAACCAACGACGCGGATTTTCTTACCTCTGGCATGCGCAATTTGGTGGGAGCGCTC 1673

QY 1172 CAGTTATTTGGAGTCCATGTCTAGACGCACTGCAAAAGTGGGAAACAGAGAGCGCTTCCATC 1231

DB 1674 CCAATTGTTGGCGTTCAITGTCGCTCGACGCAAAAGTGGGCAACAGAGCGGCTGCCACA 1733

QY 1232 CCATTGAGGAATACATGATGATACAGTTGAGAACATTTTCAGCTTCTCCAAACGCAATGA 1291

DB 1734 GTGTGGAGGAGTATATGACCTATGTTCGAGGAGTACTACCGCAGCTGGAGGTAAACGCGCA 1793

QY 1292 AAGTGGATAAAAAAGAGTGTATCTGGCCACTGATGACCTTCTTTTGTAAAGGAGGCAA 1351

Db 1794 GCACGTGTGGCTCGCGGATTTTCCTCGCATCGGATGATGCCAGGTTATTAGGAGGCGC 1853  
QY 1352 AGACAAAGTACTCCCAATTATGAATTTATTAGTGATAACTCTATTCTTTGGTCAGCTGGAC 1411  
Db 1854 GCCGAAGTACCCGAGTACCGAGATCATTGATCCGAGGTGGCGCATGGGTCGG 1913  
QY 1412 TACACAACCGATACACAGAAATTCACCTCGGGGGGTGATCCTGGATATACACTTTCTCT 1471  
Db 1914 TATCCACACGATACACGACACCGCCCTGAACGGGATCATTCTGGATATCCACCTGCTTT 1973  
QY 1472 CCCAGGCTGACTTCTTGTGTGCTACTTTTTCATCCAGGTCTGTAGGTTGCTTATGAAA 1531  
Db 1974 CCATGTCCGATCATCTGTGTGTCACCTTCTCGTCGAGGTGTGTCGCGTGGCCTACGAGA 2033  
QY 1532 TCATGCAAACTGCTGATGCTCTGCAAACTTCCATTCTTTAGATGACATCTACT 1591  
Db 2034 TAATGCAGACGATGATCCGGATGCGCATCGGTTCAAGTCGTGGACGACATATACT 2093  
QY 1592 ATTTTGGAGCCAAATGCCCAACCAAGATTGCAAGTTTATCCTCACCACCTCGAACTA 1651  
Db 2094 ACTACGGTGGCCAGAATCGCACATCGCGGTCGTTATCGCCCAAGCGCGCACGC 2153  
QY 1652 AAGAGAAATCCCATGGAACTCTGAGATATCATTTGGTGTGGCTGGAAACCATTTGGAATG 1711  
Db 2154 ACGAGGATCTACAGCTGCGTGTGTGTGACCTCGTCTCGGTGGCTGGTAAACCATTTGGGATG 2213  
QY 1712 GTTACTCTAAAGGTGTCAACAGAAACTAGGAAACAGGCTGTACCCCTTCTCTACAAAG 1771  
Db 2214 GCAATTCGAGGCAAGATACCCGACCAACCGGGGGCTATTCCTCGTTCAAGG 2273  
QY 1772 TCCGAGAGAAGATAGAAACAGTCAAAATACCCCTACATATCCTG 1813  
Db 2274 TGGAGGAGAAGGTGGACACCGCCAAAGCTGCTCTCTATGCGG 2315

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 2, 2004, 12:05:46 ; Search time 120.931 Seconds  
(without alignments)  
7328.922 Million cell updates/sec

Title: US-09-971-773-1

Perfect score: 2008

Sequence: 1 aacagaactatttctg.....gctgtgcctcaagcccatg 2008

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1568	78.1	2100	3	US-08-913-805A-9
2	1568	78.1	2100	3	US-09-442-629-9
3	1474.2	73.4	1728	3	US-08-913-805A-1
4	1474.2	73.4	1728	3	US-09-442-629-1
5	79.2	3.9	7218	1	US-08-232-463-14
6	40.8	2.0	1596	4	US-09-134-001C-2083
7	38.4	1.9	1501	4	US-09-173-300-18
8	37.6	1.9	2116	3	US-09-000-041A-1
9	37.6	1.9	5340	4	US-09-627-122-21
10	37.4	1.9	1664976	4	US-08-916-421B-1
11	37.2	1.9	4518	3	US-09-125-287-2
12	36.8	1.8	805	3	US-08-961-083-139
13	36.8	1.8	805	3	US-09-536-784-139
14	36.8	1.8	12127	4	US-08-961-527-148
15	36.2	1.8	12839	3	US-09-125-287-1
16	35.8	1.8	1664976	4	US-08-916-421B-1
17	35.6	1.8	357	4	US-09-134-001C-19
18	35.6	1.8	1368	4	US-09-134-001C-16
19	35.4	1.8	990	4	US-09-134-001C-328
20	35.2	1.8	533	1	US-08-049-283A-1
21	35.2	1.8	858	4	US-09-134-001C-1919
22	35.2	1.8	978	4	US-09-134-001C-110
23	35.2	1.8	1108	4	US-09-702-705-1249
24	35.2	1.8	1108	4	US-09-736-457-1249
25	35.2	1.8	8065	4	US-09-091-952A-6
26	35	1.7	921	4	US-09-328-352-2691
27	34.8	1.7	3487	1	US-08-410-540-4

Sequence 2, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 4, Appli  
Sequence 2, Appli  
Sequence 2383, Ap  
Sequence 12, Appli  
Sequence 3, Appli  
Sequence 1821, Ap  
Sequence 2, Appli  
Sequence 7, Appli  
Sequence 2485, Ap  
Sequence 4, Appli  
Sequence 1471, Ap  
Sequence 105, App  
Sequence 3, Appli  
Sequence 1, Appli

28 34.8 1.7 5361 3 US-08-973-462-2  
29 34.8 1.7 6152 3 US-08-973-462-1  
30 34.2 1.7 1830121 4 US-09-557-884-1  
31 34.2 1.7 1830121 4 US-09-643-990A-1  
32 34 1.7 378 1 US-08-044-547-4  
33 34 1.7 441 1 US-08-044-547-2  
34 34 1.7 699 4 US-09-134-001C-2383  
35 33.8 1.7 999 3 US-09-122-400B-12  
36 33.8 1.7 202001 4 US-09-734-674-3  
37 33.6 1.7 1260 4 US-09-134-001C-1821  
38 33.6 1.7 1835 4 US-09-512-342-2  
39 33.4 1.7 969 4 US-09-080-205-7  
40 33.4 1.7 1383 4 US-09-134-001C-2485  
41 33.4 1.7 2223 1 US-08-257-073-4  
42 33.4 1.7 3255 4 US-09-016-434-1471  
43 33.4 1.7 13158 2 US-08-687-080-105  
44 33.4 1.7 202001 4 US-09-734-674-3  
45 33.4 1.7 1230025 4 US-09-198-452A-1

#### ALIGNMENTS

#### RESULT 1

US-08-913-805A-9  
; Sequence 9, Application US/03913805A  
; Patent No. 6054304  
; GENERAL INFORMATION:  
; APPLICANT: TANIGUCHI, Naoyuki  
; APPLICANT: UOZUMI, Naofumi  
; APPLICANT: SHIBA, Tetsuo  
; APPLICANT: YANAGIDANI, Shusaku  
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3+ Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
; SOFTWARE: WordPerfect 6.1 Windows  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/913.805A  
; FILING DATE: 7 JAN 1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCY/JP97/00171  
; FILING DATE: 23 JAN 1997  
; APPLICATION NUMBER: JP 192260  
; FILING DATE: 22 JUL 1996  
; APPLICATION NUMBER: JP 162813  
; FILING DATE: 24 JUN 1996  
; APPLICATION NUMBER: JP 161648  
; FILING DATE: 21 JUN 1996  
; APPLICATION NUMBER: JP 10365  
; FILING DATE: 24 JAN 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toffenetti, Judith L.  
; REGISTRATION NUMBER: 39,048  
; REFERENCE/DOCKET NUMBER: 2356/3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-429-1776  
; TELEFAX: 202-429-0796  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2100 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double

; TOPOLOGY: linear									
; MOLECULE TYPE: cdna									
US-08-913-805A-9									
Query Match 78.1%; Score 1568; DB 3; Length 2100;									
Best Local Similarity 90.0%; Pred. No. 0;									
Matches 1691; Conservative 0; Mismatches 185; Indels 2; Gaps 1;									
Qy	1	AACAGAACTTATTTTCTGCTGTGCTACTAGAACACAGATACAAATGTTTCCAAATCTT	60						
Db	97	ACAGAGTCTATTCACCTGTGCACTAAGAACAGAGTTACAAATGTTTCAATCTT	156						
Qy	61	TCAGCTCCGAGAGACA--GAAGGAGTTGAAACTCTGAAATCGGGGATGAGCTGTT	118						
Db	157	TCAGCTCCAGGACTCCAGGGAAGTGAATTGAAATCTGAAATCGGGGATGAGCTGTT	216						
Qy	119	CTGGGCTTGAATATGCTCATCTTTTCTGCTGGGGACCTTATGTTTATATAGTG	178						
Db	217	CTGGGCTTGGATATGCTCATCTTTTCTGCTGGGGACCTTGTCTGTTTATATAGTG	276						
Qy	179	GTCAATTTGGTTTCGAGATATACCACTGACCAATCTAGCAGAGAACTCTCCAGATTC	238						
Db	277	GTCACTTGTGTAAGATATGACATCTGTATCACTTAGCCGAGAACTGTCCAGATTC	336						
Qy	239	TTGCAAGCTGGAGCGCTTAAACCAACAAATGAAGACTTGAGAGAAATGGCTGAGTCTC	298						
Db	337	TGGCAAGCTTGAACGCTTAAACAGCAGAAATGAAGACTTGAGGCGAATGGCCGAATCTC	396						
Qy	299	TCCGATACCAAGGCCCTATTTGATCAGGGGACGCTACAGGAGAGTCCGTGTTTAG	358						
Db	397	TCCGATACCAAGGCCCTATTTGATCAGGGGACGCTATAGGAGAGTACGCGTTTAG	456						
Qy	359	AAGAACAGCTTCTTAAGGCCAAGAACAGATTGAAATTTACAGAAACAAAGCTAGGAATG	418						
Db	457	AAGAGAGCTTGTTAAGGCCAAGAACAGATTGAAATTTACAGAAACAAAGCTAGGAATG	516						
Qy	419	ATCTGGGAAAGGATCATGAAATCTTTAAGGAGGAGATTGAAATCGAGCTAAAGAGTCT	478						
Db	517	GTCTGGGAAAGGATCATGAAATCTTTAAGGAGGAGATTGAAATCGAGCTAAAGAGTCT	576						
Qy	479	GGTTTTTCTACAAAGTGAATTGAAGAAATTAAGAAATTAAGAGAACAGAACTCCAA	538						
Db	577	GGTTTTTCTACAGAGTGAATTGAAGAAATTAAGAAATTAAGAGAACAGAACTCCAA	636						
Qy	539	GACATGCAGATGAAATCTTTTGGATTTAGGACATCATGAAAGGCTTATCATGACAGATC	598						
Db	637	GACATGCAGATGAAATCTTTTGGATTTAGGACATCATGAAAGGCTTATCATGACGATC	696						
Qy	599	TATACCTCAGTCAACAGATGAGCAGGATGAGTGGCGGAAAGAAAGCAGAAAGATC	658						
Db	697	TATACCTCAGTCAACAGATGAGCAGGATGAGTGGCGGAAAGAAAGCAGAAAGATC	756						
Qy	659	TCACAGCTGTCTCAGCGGAGAAATACATATCTGCAGAACTCCAGAGCTGCAGCAAG	718						
Db	757	TCACAGAACTGTCTCAGCGGAGAAATACATATCTGCAGAACTCCAGAGCTGCAGCAAG	816						
Qy	719	CCAGAAAGCTGTATGTAATATCAACAAAGGCTGTGGCTATGGCTGTCTCAGCTCATG	778						
Db	817	CCAGAAAGCTGTGTATATCAACAAAGGCTGTGGCTATGGCTGTCTCAGCTCATG	876						
Qy	779	TGGTTTATCTGCTTATGATGCTTATGCAACCGAGCAACTCATCTTGGAACTCAGA	838						
Db	877	TGGTCTACTGCTTATGATGCTTATGCAACCGAGCAACTCATCTTGGAACTCAGA	936						
Qy	839	ATTGGGCTGTCTACTGAGAGTGGGAGACTGTCTTTAGACCTGTAGTGAAGATGCA	898						
Db	937	ATTGGGCTGTCTACTGAGAGTGGGAGACTGTATTTAGGCTGTAAAGTGAAGATGCA	996						
Qy	899	CAGACAGCTGGGCTCTCCACTGGACACTGGTCAAGTGAAGTGAAGCAAAATGTTTC	958						
Db	997	CAGACAGCTGGGCTCTCCACTGGACACTGGTCAAGTGAAGTGAAGCAAAATGTTTC	1056						
Qy	959	AAGTGGTGCAGCTCCCACTTTAGACAGCCTCCATCTCTCGTCTCTTACTTACCCTTGG	1018						

Db	1057	AAGTGGTGCAGCTTCCCAATTTAGACAGCTTTCATCCCGCTCCCATATTTACCCCTGG	1116
Qy	1019	CTGTACACAGAGACTTTCAGATCGACTCTCTGAGAGTCCATGGTGCATCTCGAGTGTGGT	1078
Db	1117	CTGTACACAGAGACTTTCAGATCGACTTGTACAGTGCATGGTGCATCTCGAGTGTGGT	1176
Qy	1079	GGGTATCCAGTGTTCATAATCTTGTATCCGTCCAACTTCGCTGGCTGGAAGGGAATAG	1138
Db	1177	GGGTGTCTCAGTGTTCATAATCTTGTATCCGCCACAGCTTGGCTAGAAAAGAAATAG	1236
Qy	1139	AAGAAACCAACAGAGCTTGGCTTCAACATCCAGTTATGGAGTCCATCTCAGACGCA	1198
Db	1237	AAGAAAGCCACCAAGAGCTTGGCTTCAACATCCAGTTATGGAGTCCATCTCAGAGCGA	1296
Qy	1199	CTGACAAAGTGGGAAACAGAGCAGCTTCCATCCATTCAGTGAAGAAATACATGTCACGTTG	1258
Db	1297	CAGACAAAGTGGGAAACAGAGCTTCCATCCATTCAGTGAAGTACATGTCATGTTG	1356
Qy	1259	AAGAAACATTTTCAGCTTCTCGAACGCAAGATGAAGTGGATATAAAAAAGAGTGTATCTGG	1318
Db	1357	AAGAAACATTTTCAGCTTCTCGAACGCAAGATGAAGTGGATATAAAAAAGAGTGTATCTGG	1416
Qy	1319	CCACTGATGACCTTCTTGTAAAGGCGCAAGAGTACTCCAAATTTATGAATTTA	1378
Db	1417	CCAGATGACCTTCTTGTAAAGGCGCAAGAGTACTCCAAATTTATGAATTTA	1476
Qy	1379	TTAGTGAATACCTTATTTCTTGGTCAGCTGACATACACACCGATACACAGAAATTCAC	1438
Db	1477	TTAGTGAATACCTTATTTCTTGGTCAGCTGACATACACACCGATACACAGAAATTCAC	1536
Qy	1439	TTGGGGCGTGAATCTCGATATACATTTCTCTCCAGGCTGACTTCTTGTGTGTTACTT	1498
Db	1537	TTCTGGAGTGAATCTCGATATACATTTCTCTCTCAGGCGAGCTTCTTGTGTGTTACTT	1596
Qy	1499	TTTCATCCAGCTGTAGGGTGTCTTATGAAATCATGCAACACACTGCATCTCTGCTCT	1558
Db	1597	TTTCATCCAGCTGTAGGGTGTCTTATGAAATCATGCAACACACTGCATCTCTGCTCT	1656
Qy	1559	CTGCAAACTTCCATCTTTTAGATGACATCTACTATTTTGGAGGCGCAAAATGCCACAACC	1618
Db	1657	CTGCAAACTTCCATCTTTTAGATGACATCTACTATTTTGGGGCGCAGATGCCCAATC	1716
Qy	1619	AGATTGAGTGTATCTCAGCAACCTCGAACTAAAGAGGAAATCCCATGGAACCTGGAG	1678
Db	1717	AAATTGCAATTTATGCTCACCACCCGCACTGCAGATGAAATTTCCCATGGAACCTGGAG	1776
Qy	1679	ATATCATTTGGTGGCTGGAAACCAATTTGGAATGGTACTCTTAAAGGTCTCAACAGAAAC	1738
Db	1777	ATATCATTTGGTGGCTGGAAATCAATTTGGAGTGGCTATTTCTTAAAGGTCTCAACAGAAAT	1836
Qy	1739	TAGAAAAACAGGCTGTACCTTCTCACAAGTCCGAGAGAGATAGAAACAGTCAAAAT	1798
Db	1837	TGGGAGGACGGGCTTATATCCCTCTTCAAAAGTTCCAGAGAGATAGAAACGCTCAAGT	1896
Qy	1799	ACCTCATATCTCTGAAAGCTGAAAAATAGAGATGGAGTGAAGAGATTAAACACAGAAAT	1858
Db	1897	ACCCCATATCTCTGAGGCTGAGAAATAAAGCTCAGATGGAAGAGATAAACGACCAAACT	1956
Qy	1859	TAGTTACAGCACTCTCAG	1876
Db	1957	CAGTTGACCAAACTCAG	1974

RESULT 2  
US-09-442-629-9  
; Sequence 9, Application US/09442629  
; Patent No. 6291219  
; GENERAL INFORMATION:  
; APPLICANT: TANIGUCHI, Naoyuki  
; UOJUMI, Naofumi  
; SHIBA, Tetsuo  
; YANAGIDANI, Shusaku



1439 TTTCGGGCGTGTATCTCGGATATACACTTTCTCTCCAGGCTGACTTCCTTGTGTACTT 1498  
1537 TTTCGGGAGTGTATCTCGGATATACACTTTCTCTCTCAGGAGACTTCCTAGTGTACTT 1596  
1499 TTTCATCCAGGCTGTAGGTTGCTTATGAATCATGCAACATGCTGATGCT 1558  
1597 TTTCATCCAGGCTGTAGGTTGCTTATGAATCATGCAACATGCTGATGCT 1656  
1559 CTGCAAACTTTTATCTTTAGATGACATCTACTATTTTGGAGGCCAAATGCCACAACC 1618  
1657 CTGCAAACTTTTATCTTTAGATGACATCTACTATTTTGGAGGCCAAATGCCACAATC 1716  
1619 AGATTGCACTTTTATCTCTCAACCTCGAACTTAAAGAGAAATCCCATGGAACCTGGAG 1678  
1717 AAATTGCCATTTATGCTCAACCAACCCGAACTGCAGATGAAATTCCTATGGAACCTGGAG 1776  
1679 ATATCATTTGTTGCTGGAACCAATGGAATGGTTACTCTTAAAGTGTCAACAGAAAC 1738  
1777 ATATCATTTGTTGCTGGAATCATTTGGGATGGCTATTTCTAAAGTGTCAACAGAAAT 1836  
1739 TAGGAAACACGCTGTACCTTCTCAAAAGTCCGAGAGAGATAGAAACAGTCAAT 1798  
1837 TGGGAAGGACGGGCTTATATCTCTCAAAAGTTCGAGAGAGATAGAAACGTTCAAGT 1896  
1799 ACCCTACATATCTCTGAAGCTGAAATATAGAGATGGAGTGTAGAGATTTAAACACAGAAAT 1858  
1897 ACCCACATATCTCTGAGGCTGAGAAATAAAGCTCAGATGGAAGATAGAAACGCAAACT 1956  
1859 TAGTTGAGACCATCTCAG 1876  
1957 CAGTTGAGCAAACTCAG 1974

## RESULT 3

US-08-913-805A-1  
Sequence 1, Application US/08913805A  
Patent No. 6054304  
GENERAL INFORMATION:  
APPLICANT: TANIGUCHI, Naoyuki  
APPLICANT: UOZUMI, Naofumi  
APPLICANT: SHIBA, Tetsuo  
APPLICANT: YANAGIDANI, Shusaku  
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenyon & Kenyon  
STREET: 1025 Connecticut Avenue, N.W., Suite 600  
CITY: Washington  
STATE: DC  
COUNTRY: US  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3+ Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/913,805A  
FILING DATE: 7 JAN 1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP97/00171  
FILING DATE: 23 JAN 1997  
APPLICATION NUMBER: JP 192260  
FILING DATE: 22 JUL 1996  
APPLICATION NUMBER: JP 162813  
FILING DATE: 24 JUN 1996  
APPLICATION NUMBER: JP 161648  
FILING DATE: 21 JUN 1996  
APPLICATION NUMBER: JP 10365  
FILING DATE: 24 JAN 1996  
ATTORNEY/AGENT INFORMATION:

NAME: Toffenetti, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 2356/3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1728 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-913-805A-1

Query Match 73.4%; Score 1474.2; DB 3; Length 1728;

Best Local Similarity 90.9%; Pred. No. 0;

Matches 1569; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

Qy	100	ATGCGGCATGGACTGGTTCCTGGCGTTGGATTATGCTCAATCTTTTTCCTGGGGGACC	159
Db	1	ATGCGGCATGGACTGGTTCCTGGCGTTGGATTATGCTCAATCTTTTTCCTGGGGGACC	60
Qy	160	TTATTGTTTATATAGTGGTCAATTTGGTTCGAGATATGACCACTTCTAGC	219
Db	61	TTGCTATTTTACATAGTGGTCAATTTGGTTCGAGATATGACCACTTCTAGC	120
Qy	220	AGAGAACTCTCCAAAGATTCTTGAAAGCTGGAGCGCTTAAACAAACAAATGAAGACTTG	279
Db	121	CGAGAACTGTCCAAAGATTTTGGCAAGCTGGAGCGCTTAAACAAACAAATGAAGACTTG	180
Qy	280	AGGAAATGGCTGAGTCTCTCCGAATACAGAAAGCCCTTATGATCAGGGGACAGCTACA	339
Db	181	AGGAAATGGCTGAGTCTCTCCGAATACAGAAAGCCCTTATGATCAGGGGACAGCTACA	240
Qy	340	GGAAGAGTCGGTGTTTAGAGAAACAGCTTGTAAAGCCCAAGACAGATGAAATTTAC	399
Db	241	GGAAGAGTCGGTGTTTAGAGAAACAGCTTGTAAAGCCCAAGACAGATGAAATTTAT	300
Qy	400	AAGAAACAAAGCTAGGAATGATCTGGGAAAGGATCATGAAATCTTAAAGGAGGAGATTGAA	459
Db	301	AAGAAACAAAGCTAGGAATGATCTGGGAAAGGATCATGAAATCTTAAAGGAGGAGATTGAA	360
Qy	460	AATGAGCTAAAGAGCTCTGGTCTTCTCAAAAGTGAATTTGAAGAAATTTAAAGAAATTA	519
Db	361	AATGAGCTAAAGAGCTCTGGTCTTCTCAAAAGTGAATTTGAAGAAATTTAAAGAAATTA	420
Qy	520	GAAGAAACGAATCCCAAGACATGAGATGAAATCTTTTGGATTTAGGACATCATGAA	579
Db	421	GAAGAAATGAATCCCAAGACATGAGATGAAATCTTTTATCAGATTTGGGACATCATGAA	480
Qy	580	AGGTCTATCATGACAGATCTATCTACTCTCAGTCAAAACAGATGGAGAGTGGTGGCG	639
Db	481	AGGTCTATCATGACAGATCTATCTACTCTCAGTCAAAACAGATGGGAGGAGTGGCGT	540
Qy	640	GAAGAAAGCAAGCTTGAAGAGCTGGTCCAGCGGAGATAACATCTCCAGAT	699
Db	541	GAAGAAAGCAAGCTTGAAGAGCTGGTCCAGCGGAGATAACATCTCTCAGAT	600
Qy	700	CCAGAGACTGCAGCAAGCAAGCTGGTATGTAATATCAACAAAGCTGTGGCTAT	759
Db	601	CCAGAGACTGCAGCAAGCAAGCTGGTATGTAATATCAACAAAGCTGTGGCTAT	660
Qy	760	GGATGTCAACTCCATCATGTGGTGTACTGTCTTATGTTATGTTATGGCACCAGCAACA	819
Db	661	GGCTGTCAAGTCCATCATGTAGTGTACTGTCTTATGTTATGTTATGGCACCAGCAACA	720
Qy	820	CTCATCTTGAATCTCAGAAATGGCGCTATGCTACTCGAGATGGGAGACTGTGTTAGA	879
Db	721	CTGCGCTTGGAACTCTCAAAATGGCGCTACTCGGAGATGGGAAACTGTGTTAGA	780
Qy	880	CTGTAAAGTGAACATGCACAGAGCTCTGGCTCTCCACTGGACACTGTCAGGTGAA	939

Db 781 CCTGTAGTGCAGCGTGCACAGACAGATCTGGCAGCTCCACTGGACATTTGGTCAAGTGAA 840  
Qy 940 GTGAAGGACAAAATGTTCAAGTGTGTGAGTCTCCCATTTGTAGACAGCTCCATCCTCGT 999  
Db 841 GTAAAGGACAAAATGTTCAAGTGTGTGAGTCTCCCATTTGTAGACAGTGTTCATCCTCGT 900  
Qy 1000 CCTCTTACTTACCCCTTGGCTGTACCAAGAACCTTTGCAGATCGACTCCTGAGAGTCCAT 1059  
Db 901 CCTCATATTTACCCCTTGGCTGTCCCAAGAACCTTTGCAGATCGACTTGTACGAGTCCAT 960  
Qy 1060 GGTGATCTCAGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1119  
Db 961 GGTGATCTCAGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1020  
Qy 1120 TGGCTGGAAGGGAATAAGAAACACCAAGAGCTTTGGCTTCAAAACATCCAGTTATT 1179  
Db 1021 TGGCTGGAAGGGAATAAGAAACACCAAGAGCTTTGGCTTCAAAACATCCAGTTATT 1080  
Qy 1180 GGAGTCCATGTGACAGCACTGACAAAGTGGAAACAGAGCGCTTCCATCCCATTTGAG 1239  
Db 1081 GGAGTCCATGTGACAGCACTGACAAAGTGGAAACAGAGCGCTTCCATCCCATTTGAG 1140  
Qy 1240 GAATACATGTTACAGTGTGAGAACATTTTCAGCTTCTCGAACGAGAAATGAAGTGGAT 1299  
Db 1141 GAATACATGTTACAGTGTGAGAACATTTTCAGCTTCTCGAACGAGAAATGAAGTGGAT 1200  
Qy 1300 AAAAAAGAGTGTATCTGCCCACTGATGACCTCTCTTTTAAAGGAGCAAGACAAAG 1359  
Db 1201 AAAAAAGAGTGTATTTGGCCACAGATGACCTCTCTTTTAAAGGAGCAAGACAAAG 1260  
Qy 1360 TACTCCAATTATGAATTTTATGATGAATCTATTTCTTGGTCACTGGAGTACACAAAC 1419  
Db 1261 TACCCAGTTATGAATTTTATGATGAATCTATTTCTTGGTCACTGGAGTACATAAT 1320  
Qy 1420 CGATACACAGAAATTCATCTCGGGCGTGTATCTCGATATACACTTTCTCTCCAGGCT 1479  
Db 1321 CGATATACAGAAATTCATCTCGGGCGTGTATCTCGATATACACTTTCTCTCCAGGCA 1380  
Qy 1480 GACTTCTTGTGTGTACTTTTTCATCCAGGCTGTAGGTTGCTTATGAATCATGCAA 1539  
Db 1381 GACTTCTTGTGTGTACTTTTTCATCCAGGCTGTAGGTTGCTTATGAATCATGCAA 1440  
Qy 1540 ACATGTCATCTGATGCTCTGCAAACTTCCATTTCTTTAGATGACATCTACTATTTTGA 1599  
Db 1441 GCGTGCATCTGATGCTCTGCAAACTTCCATTTCTTTAGATGACATCTACTATTTTGA 1500  
Qy 1600 GGCCAAATGCCCAACACAGATTCAGTGTATCTCAACCACTCGAACTTAAAGAGGAA 1659  
Db 1501 GGCCAAATGCCCAACACCAAAATGCCCAATTTATCTCAACCACTCGAACTTAAAGAGGAA 1560  
Qy 1660 ATCCCATGGAACTGGAGATATCATTTGGTGGCTGGAAACCATTTGAAATGTTACTCT 1719  
Db 1561 ATCCCATGGAACTGGAGATATTAATTTGGTGGCTGGAAATCACTGGGATGGCTATCTCT 1620  
Qy 1720 AAAGTGTCAACAGAAAACTAGGAAAAACAGGCTGTACCCCTTCTCAAAAGTCCGAGAG 1779  
Db 1621 AAAGTGTCAACAGAAAACTGGAGAGCGGCGCTATATCTCTCAAAAGTCCGAGAG 1680  
Qy 1780 AAGATAGAAAAGTCAAAATACCTACATATCTCTGAAGCTGAAAAATA 1826  
Db 1681 AAGATAGAAAAGTCAAAATACCTACATATCTCTGAAGCTGAAAAATA 1727

## RESULT 4

US-09-442-629-1

; Sequence 1, Application US/09442629

; Patent No. 6291219

; GENERAL INFORMATION:

; APPLICANT: TANIGUCHI, Naoyuki

; UOZUMI, Naofumi

; SHIBA, Tetsuo

; YANAGIDANI, Shusaku

; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase

NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Kenyon &amp; Kenyon

STREET: 1025 Connecticut Avenue, N.W., Suite 600

CITY: Washington

STATE: DC

COUNTRY: US

ZIP: 20036

COMPUTER READABLE FORM:

MEDIUM TYPE: 3+ Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS 6.2

SOFTWARE: WordPerfect 6.1 Windows

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/442.629

FILING DATE: 18-No. 6291219-1999

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/913.805A

FILING DATE: 7 JAN 1998

APPLICATION NUMBER: PCT/JP97/00171

FILING DATE: 23 JAN 1997

APPLICATION NUMBER: JP 192260

FILING DATE: 22 JUL 1996

APPLICATION NUMBER: JP 162813

FILING DATE: 24 JUN 1996

APPLICATION NUMBER: JP 161648

FILING DATE: 21 JUN 1996

APPLICATION NUMBER: JP 10365

FILING DATE: 24 JAN 1996

ATTORNEY/AGENT INFORMATION:

NAME: Toffenetti, Judith L.

REGISTRATION NUMBER: 39,048

REFERENCE/DOCKET NUMBER: 2356/3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-429-1776

TELEFAX: 202-425-0796

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1728 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-442-629-1

Query Match 73.4%; Score 1474.2; DB 3; Length 1728;  
Best Local Similarity 90.9%; Pred. No. 0;  
Matches 1569; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

Qy 100 ATGCGGCGCATGGACTGTTCTCTGGCGTTGGATTATGCTCATTTCTTTTGGCTGGGGGACC 159  
Db 1 ATGCGGCGCATGGACTGTTCTCTGGCGTTGGATTATGCTCATTTCTTTTGGCTGGGGGACC 60  
Qy 160 TTATTTGTTTTATAGTGTGTTCTTGGTTCGAGATATGACCACTTGTAGCATTTCTAGC 219  
Db 61 TTGCTATTTTACATAGTGTGTTCTTGGTTCGAGATATGACCACTTGTAGCATTTCTAGC 120  
Qy 220 AGAGAACTCTCCAAAGATCTTTGCAAGCTGGAGCGCTTTAAACAAACAAATGAAGACTTG 279  
Db 121 CGAGAACTGTCCAAAGATTTTGGCAAGCTGGAGCGCTTTAAACAAACAAATGAAGACTTG 180  
Qy 280 AGGAGATGGCTGAGTCTCTCCGAATACCAAGGCGCTTATTTGATCAGGGGACGACTACA 339  
Db 181 AGGAGATGGCTGAGTCTCTCCGAATACCAAGGCGCTTATTTGATCAGGGGACGACTTCA 240  
Qy 340 GGAAGAGTCCGTGTTTGTAGAGAAACAGCTTCTTAAGGCCAAAGAACAGATTGAAAAATTAC 399  
Db 241 GGAAGAGTCCGTGTTTGTAGAGAAACAGCTTCTTAAGGCCAAAGAACAGATTGAAAAATTAT 300  
Qy 400 AAGAAACAAAGCTAGGAATGATCTGGGAAAGGATCATGAAATCTTTAAGGAGGAGATTGAA 459  
Db 301 AAGAAACAAACTAAAAATGGTCCAGGGAAGGATCATGAAATCTTTAAGGAGGAGATTGAA 360

QY 460 AATGAGCTAAGAGCTCTGTTTTTTCTACAAAGTGAATTAAGAAATTAAGAAATTA 519  
Db 361 AATGAGCTAAGAGCTCTGTTTTTTCTACAAAGTGAATTAAGAAATTAAGAAATTA 420  
QY 520 GAAGGAACGAGCTCAAGAGCATGAGTGAATTTCTTTGGATTTAGGACATCATGAA 579  
Db 421 GAAGGAATGAATCTCAAGAGCATGAGTGAATTTCTTTGGATTTAGGACATCATGAA 480  
QY 580 AGGTCTATCATGACAGCTCTTACTACCTCAGTCAACAGAGTGGAGCAGGTGAGTGGCGG 639  
Db 481 AGGTCTATTAAGAGCATGATCTATCTACTCAGTCAACAGAGTGGAGCAGGTGAGTGGCGT 540  
QY 640 GAAAGAGAGCCAAAGATCTGACAGAGCTGGTCCAGCGGAGATTAACATATCTGAGAAAT 699  
Db 541 GAAAGAGAGCCAAAGATCTGACAGAGCTGGTCCAGCGGAGATTAACATATCTTTCAGAAAT 600  
QY 700 CCAAGAGCTGAGCAAGAGCCAAAGAGCTGGTATGATTAATATCAACAAAGGCTGGCTAT 759  
Db 601 CCAAGAGCTGAGCAAGAGCCAAAGAGCTGGTATGATTAATATCAACAAAGGCTGGCTAT 660  
QY 760 GGATGCTCAACTCCATCATGTTGTTTACTGCTTCTCATGATTGTCGACCCAGCGAACA 819  
Db 661 GGCTGTCAGCTCCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
QY 820 CTGATCTTTGGAATCTCAGAAATTTGGCGCTATGCTACTGGAGGATGGAGACTGTGTTTGA 879  
Db 721 CTGCGCTTTGGAATCTCAGAAATTTGGCGCTACTGCTACTGGGAGTGGAAACTGTGTTTGA 780  
QY 880 CTTGTAAGTGAATGATGACAGAGCTGGCTCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCT 939  
Db 781 CTTGTAAGTGAATGATGACAGAGCTGGCTCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
QY 940 GTGAGGACAAATGTTTCAAGTGTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 999  
Db 841 GTGAGGACAAATGTTTCAAGTGTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
QY 1000 CTTCTTACTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1059  
Db 901 CTTCTTACTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960  
QY 1060 GGTGATCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1119  
Db 961 GGTGATCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
QY 1120 TGGCTGGAAGGGAATAGAGAAACCAAGAGCTTGGCTTCAAAATCCAGTTATT 1179  
Db 1021 TGGCTGGAAGGGAATAGAGAAACCAAGAGCTTGGCTTCAAAATCCAGTTATT 1080  
QY 1180 GGAGTCCATGTCAGAGCAGTGAAGAGTGGGAGCAGAGCAGTTCCTCCATCCATTCGAG 1239  
Db 1081 GGAGTCCATGTCAGAGCAGTGAAGAGTGGGAGCAGAGCAGTTCCTCCATCCATTCGAG 1140  
QY 1240 GAATACATGTTGACGTTGAAGAAATTTTTCAGCTTCTGAAAGCAGAGTGAAGTGGAT 1299  
Db 1141 GAATACATGTTGACGTTGAAGAAATTTTTCAGCTTCTGAAAGCAGAGTGAAGTGGAT 1200  
QY 1300 AAAAAAGAGTGTATCTGCGCACTCATGACCTTCTTTGTTAAAGAGGCAAGAGCAAG 1359  
Db 1201 AAAAAAGAGTGTATCTGCGCACTCATGACCTTCTTTGTTAAAGAGGCAAGAGCAAG 1260  
QY 1360 TACTCCAAATTTATGATTTATTTAGTGAATCTATTTCTGCTCAGCTGGAGTCAACAA 1419  
Db 1261 TACTCCAAATTTATGATTTATTTAGTGAATCTATTTCTGCTCAGCTGGAGTCAACAA 1320  
QY 1420 CGATACAGAAAAATTTCTGCGGCGTGTATCTGATGATATACATTTCTCTCCAGCT 1479  
Db 1321 CGATATACAGAAAAATTTCTGCGGCGTGTATCTGATGATATACATTTCTCTCCAGCT 1380  
QY 1480 GACTTCTTGTGTGTTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1539  
Db 1381 GACTTCTTGTGTGTTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440

QY 1540 AACTGTCATCTGATGCTCTGCAAACTTCCATCTTTAGATGACATCTATTTTGA 1599  
Db 1441 GCGTGTGATCTGATGCTCTGCAAACTTCCATCTTTAGATGACATCTATTTTGA 1500  
QY 1600 GGCCAAATGCCCAACACAGATTCGAGTTTATCTCCTCACCAACCTCGAACTAAAGAGGAA 1659  
Db 1501 GGCCAAATGCCCAACACAGATTCGAGTTTATCTCCTCACCAACCTCGAACTAAAGAGGAA 1560  
QY 1660 ATCCCCATGGAACTGGAGATATCATTTGGTGTGGTGGGAAACCATTTGGAATGGTTACTCT 1719  
Db 1561 ATCCCCATGGAACTGGAGATATCATTTGGTGTGGTGGGAAACCATTTGGAATGGTTACTCT 1620  
QY 1720 AAAGTGTCAACAGAACTAGGAAACACAGGCTGTACCTTCTTACAAAGTCCGAGAG 1779  
Db 1621 AAAGTGTCAACAGAACTAGGAAACACAGGCTGTACCTTCTTACAAAGTCCGAGAG 1680  
QY 1780 AAGATAGAAACAGTCAAAATACCTTACATATCTTGAAGCTGAAAAATA 1826  
Db 1681 AAGATAGAAACAGTCAAAATACCTTACATATCTTGAAGCTGAAAAATA 1727

## RESULT 5

US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-F18  
; US-08-232-463-14

Query Match 3.9%; Score 79.2; DB 1; Length 7218;  
Best Local Similarity 6.6%; Pred. No. 1e-12;  
Matches 27; Conservative 236; Mismatches 149; Indels 0; Gaps 0;





## US-09-000-041A-1

Query Match 1.9%; Score 37.6; DB 3; Length 2116;  
Best Local Similarity 49.6%; Pred. No. 0.75;  
Matches 126; Conservative 0; Mismatches 124; Indels 4; Gaps 1;

QY 379 AAAGAACAGATTGAAATTTACAGAAACAAAGCTAGGAATGATCTCGGAAAGATCATGAA 438  
DB 1507 AATGAAACAAATGGAGCCCTGGCTATCCCAAGCCAATCAAGTCTCGAAAGGATCCCTGA 1566

QY 439 ATCTTAAGGAGCAGATTGAAATCGAGCTAAAGAGCTCTGTTTTTTCTACAAG---- 494  
DB 1567 ATCTCCTCAGGAGGATTGTACACAAAGAAATGCGTTTACGTTATTTCTCAAGAGGAA 1626

QY 495 TGAATTGAAGAAATTAAGAAATTAAGAGGAAACCAACTCCAAAGACATCGAGATGAAAT 554  
DB 1627 GGAGTATTGGAATTTCAACACCAAGATCTCAAGTAGAACCTGACATCCAAGATCAT 1686

QY 555 TCTTTTGGATTAGACATCATGAAGGTCTATCATGACAGATCTATCTACCTCAGTCA 614  
DB 1687 CCTCAAGATTTTATGGGCTGTGATGGACCAACAGACAGAGTTAAAGAGGACACAGCCC 1746

QY 615 AACAGATGAGCAG 628  
DB 1747 ACCAGATGATGAG 1760

## RESULT 9

US-09-627-122-21/c  
; Sequence 21, Application US/09627122  
; Patent No. 6472521  
; GENERAL INFORMATION:  
; APPLICANT: Uhlmann, Eugen  
; APPLICANT: Greiner, Beate  
; APPLICANT: Unger, Eberhard  
; APPLICANT: Gothe, Gislinde  
; APPLICANT: Schwerdel, Marc  
; TITLE OF INVENTION: OLIGONUCLEOTIDES FOR THE INHIBITION OF HUMAN EGS  
; FILE REFERENCE: 02481.1678  
; CURRENT APPLICATION NUMBER: US/09/627,122  
; CURRENT FILING DATE: 2000-07-27  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 21  
; LENGTH: 5340  
; TYPE: DNA  
; ORGANISM: Plasmodium falciparum  
US-09-627-122-21

Query Match 1.9%; Score 37.6; DB 4; Length 5340;  
Best Local Similarity 54.3%; Pred. No. 1.2;  
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 426 AAAGATCATGAAATCTTAAAGGAGGAGGATGAAATGGAGCTAAAGAGCTCTGTTTTT 485  
DB 3851 AAAAGAGCAGCAGCTGAAACGCTTATAATAAAAAATCGAGTAGAAGTCATTGATTTT 3792

QY 486 TCTCAAGTGAATTTGAAGAAATTAAGAAATTAAGAGGAAACGAACTCCAAAGCATGC 545  
DB 3791 ACTATAACATTAATTTATAAAGATATAAATATGTAGGAGAAAGTATAACCAAAATAGGA 3732

QY 546 AGATGAAATCTTTTGGATT 565  
DB 3731 AAATTAATTTAGTTGATT 3712

## RESULT 10

US-08-916-421B-1/c  
; Sequence 1, Application US/08916421B  
; Patent No. 6503729  
; GENERAL INFORMATION:  
; APPLICANT: Bult et al.

; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus  
; Patent No. 6503729  
; TITLE OF INVENTION: jannaschii  
; FILE REFERENCE: PB275  
; CURRENT APPLICATION NUMBER: US/08/916,421B  
; CURRENT FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: US 60/024,428  
; PRIOR FILING DATE: 1996-08-22  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 1664976  
; TYPE: DNA  
; ORGANISM: Methanococcus jannaschii  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (28222)..(28222)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (28257)..(28258)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (84773)..(84773)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (84808)..(84808)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (84812)..(84812)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98120)..(98120)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98159)..(98159)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98239)..(98239)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98266)..(98266)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98343)..(98343)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (103998)..(103998)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (148948)..(148948)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (163385)..(163385)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (191989)..(191989)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (191995)..(191995)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (231980)..(231980)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (234187)..(234187)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (234220)..(234220)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (234814)..(234814)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature

```

; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664855)
; OTHER INFORMATION: n equals a, t, c, or g
;
US-08-916-421B-1

Query Match          1.4%; Score 37.4; DB 4; Length 1664976;
Best Local Similarity 47.3%; Pred. No. 22;
Matches 113; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY 351 TGTTTTAGAAGACAGCTTGTTAGGCCAAGAACAGATTGAAATTTACAGAAACAAGC 410
    |||||
Db 331793 TGTCTTAAGAGCACCAATTGGATGGTATAGGCATTTAAAAATCAGCTGTAAAGGCATCC 331734
    |||||

QY 411 TAGGAATGATCTGGGAAGGATCATGAAATCTTAAAGGAGGAGGATGAAATGGAGCTAA 470
    |||||
Db 331733 TTTTGAGTGAGTTGTTCAGAAAGATTGTAGCTAAGGAAAGAAAGAAAGAGGAGAAGA 331674
    |||||

QY 471 AGAGCTCTGGTTTTTTCTTCAAAAGTGAATTTGAAGAAATTTAGAAGGAAAGCA 530
    |||||
Db 331673 ATCAAAATTTCTACTTATTAACCCAGAAACAGAGGAAATTTTGAATTTGAATGAAACAA 331614
    |||||

QY 531 ACTCCAAGACATGCAGATGAAATTCCTTTGGATTTTAGGCATCATGAAAGCTTATCA 589
    |||||
Db 331613 CATAAATATTAATTAATCAATCAAGAGATTGTTAGCTTTAGCTTAAACATGAATTTGGGAATTA 331555
    |||||

```

US-09-125-287-2/c

; Patent No. 6114602

APPLICANT: BARG, Ri

**TITLE OF INVENTION: MET**

; FILE REFERENCE: INTRO GENETICS

; CURRENT FILING DATE: 1998-11-11

; EARLIER FILING DATE: 1997-0

; SOFTWARE: PatentIn Ver. 2.0

; LENGTH: 4518

ORGANISM: TPRP-F1 PROMOTOR

```

; NAME/KEY: unsure
; LOCATION: (C) (101)

```

OTHER INFORMATION: IIS-09-125-287-2



APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,527  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 148:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12127 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-148

Query Match 1.8%; Score 36.8; DB 4; Length 12127;  
Best Local Similarity 50.6%; Pred No. 3;  
Matches 89; Conservative 0; Mismatches 87; Indels 0; Gaps 0;  
Qy 272 AAGACTGAGGAGGATGCTGAGTCTCCGAATACCAAGAGGCGCTTATCATCAGGGGA 331  
Db 8181 AAGAAATTTACAAACAAATTTATCAAAAAGTTGATGAAACTTTTAAACAAGAGA 8240  
Qy 332 CAGCTACGAGGAGAGTCGCTGTTTGAAGAACAGCTTTGTAAGCCCAAGAACAGATTG 391  
Db 8241 ATTTGGAAGAACTCTTAAATCTCTAAATGATCTTTGTTGATAAATATCAAAAACAAATCG 8300  
Qy 392 AAAATTACAGAAACAACTAGGATGATCTGGGAAAGGATCATGAAATCTTAAGG 447  
Db 8301 AACTTTGAGAAAGAGAGAGAAAAGCTGCTGAAAAAGCTGCTGAAAAAGCAAG 8356

RESULT 15  
US-09-125-287-1/c  
; Sequence 1, Application US/09125287B  
; Patent No. 6114602  
; GENERAL INFORMATION:  
; APPLICANT: BARG, Rivka  
; APPLICANT: SALT, Tehiam  
; TITLE OF INVENTION: METHOD FOR THE INTRODUCTION OF GENETIC PARTHENOCAOPY IN  
; TITLE OF INVENTION: PLANTS  
; FILE REFERENCE: INTRON GENETIC PARTHENOCAOPY IN PLANTS  
; CURRENT APPLICATION NUMBER: US/09/125,287B  
; CURRENT FILING DATE: 1998-11-09  
; EARLIER APPLICATION NUMBER: PCT/IL97/00051  
; EARLIER FILING DATE: 1997-02-13  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 12839  
; TYPE: DNA

ORGANISM: TPRP-F1 GENOMIC CLONE  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (5) (11322)  
OTHER INFORMATION: "n"s are any nucleic residue  
US-09-125-287-1  
Query Match 1.8%; Score 36.2; DB 3; Length 12839;  
Best Local Similarity 22.5%; Pred. No. 4.6;  
Matches 49; Conservative 67; Mismatches 102; Indels 0; Gaps 0;  
Qy 1586 TCTACTATTTTGGAGGCCAAATCCCAACACAGATTGCGATTATCTTACCAACCTC 1645  
Db 218 TCGCTATTTGGGATGCATTATTHDSVANAASNATANNAYGSNTYADASNASHVHYN 159  
Qy 1646 GAACATAAGAGGAATCCCATGGAACCTGGAGATATCATTTGGTGGCTGGAACCAT 1705  
Db 158 ADASRTNTNSYYAKYSACDAYHSANHNATANNATNTYADASNASSYAAHVNHN 99  
Qy 1706 GGAATGGTTACTCTAAAGGTGTCAACAGAAACTAGGAAACACAGGCTGTACCTCTCT 1765  
Db 98 HGCATDAYYTHDSVASTWKYCNHGNAGNASDAYYTHDSVAANSTWKYCNHNANCYAKYDA 39  
Qy 1766 ACAAGTCCGAGAGAGATAGAAACAGTCAATACCT 1803  
Db 38 GNASDANCYAKYDAVNCNHAGNNGKNCYADAGNASVT 1  
Search completed: February 2, 2004, 18:41:04  
Job time : 152.431 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 2, 2004, 12:16:51 ; Search time 702.477 Seconds  
(without alignments)  
10417.799 Million cell updates/sec

Title: US-09-971-773-1  
Perfect score: 2008  
Sequence: 1 aacagaaactatttctg.....gctgtgcctcaagcccatg 2008

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2434939 seqs, 1822278265 residues

Total number of hits satisfying chosen parameters: 4869878

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US05\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2008	100.0	2008	11 US-09-971-773-1	Sequence 1, Appli
2	1584	78.9	1728	11 US-09-971-773-2	Sequence 2, Appli
3	1574.2	78.4	3007	15 US-10-106-698-381	Sequence 381, App
4	1569.6	78.2	2100	9 US-09-839-136-9	Sequence 9, Appli
5	1472.6	73.3	1728	9 US-09-839-136-1	Sequence 1, Appli
6	976.4	48.6	979	11 US-09-971-773-6	Sequence 6, Appli
7	893.2	44.5	422	11 US-09-971-773-7	Sequence 7, Appli
8	368.2	18.3	979	11 US-09-918-995-33468	Sequence 33468, A
9	302	15.0	9196	11 US-09-971-773-3	Sequence 3, Appli
10	285.6	14.2	503	9 US-09-864-761-13359	Sequence 13359, A
11	218.8	10.9	551	9 US-09-864-761-13292	Sequence 13292, A
12	213.4	10.6	248	9 US-09-864-761-18181	Sequence 18181, A
13	213.4	10.6	248	9 US-09-864-761-30284	Sequence 30284, A
14	149	7.4	481	13 US-10-125-968-274	Sequence 274, App
15	136.4	6.8	384	9 US-09-864-761-1421	Sequence 1421, Ap

16	117.4	5.8	151	12	US-10-242-535A-52289	Sequence 52289, A
17	112.8	5.6	419	11	US-09-918-995-33106	Sequence 33106, A
18	103.4	5.1	129	9	US-09-864-761-29920	Sequence 29920, A
19	53.8	2.7	65	13	US-09-908-975-23971	Sequence 23971, A
20	46.2	2.3	60	13	US-09-908-975-4992	Sequence 4992, Ap
21	42.8	2.1	1137	15	US-10-184-644-518	Sequence 518, App
22	42.8	2.1	1137	15	US-10-184-634-518	Sequence 518, App
23	40.8	2.0	3673778	13	US-10-312-841-2	Sequence 2, Appli
24	39	1.9	581	10	US-09-789-054A-11	Sequence 11, Appli
25	38.8	1.9	1485	10	US-09-764-877-3640	Sequence 3640, Ap
26	38.8	1.9	1485	11	US-09-764-891-5860	Sequence 5860, Ap
27	38.8	1.9	1485	12	US-10-242-515-3640	Sequence 3640, Ap
28	38.4	1.9	1501	14	US-10-027-450-18	Sequence 18, Appli
29	38.2	1.9	11872	10	US-09-764-847-1390	Sequence 1390, Ap
30	38.2	1.9	11872	15	US-10-092-154-1390	Sequence 1390, Ap
31	38.2	1.9	3673778	13	US-10-312-841-1	Sequence 1, Appli
32	38	1.9	716	13	US-10-140-472-96	Sequence 96, Appli
33	38	1.9	716	13	US-10-141-761-96	Sequence 96, Appli
34	38	1.9	716	13	US-10-142-885-96	Sequence 96, Appli
35	38	1.9	716	13	US-10-158-790-96	Sequence 96, Appli
36	38	1.9	716	13	US-10-137-871-96	Sequence 96, Appli
37	38	1.9	716	13	US-10-140-805-96	Sequence 96, Appli
38	38	1.9	716	13	US-10-140-864-96	Sequence 96, Appli
39	38	1.9	716	13	US-10-140-923-96	Sequence 96, Appli
40	38	1.9	716	13	US-10-141-756-96	Sequence 96, Appli
41	38	1.9	716	13	US-10-141-759-96	Sequence 96, Appli
42	38	1.9	716	15	US-10-123-155-96	Sequence 96, Appli
43	38	1.9	716	16	US-10-146-731-96	Sequence 96, Appli
44	38	1.9	941	13	US-10-140-472-464	Sequence 464, App
45	38	1.9	941	13	US-10-141-761-464	Sequence 464, App

## ALIGNMENTS

## RESULT 1

US-09-971-773-1  
; Sequence 1, Application US/09971773  
; Publication No. US20030115614A1  
; GENERAL INFORMATION:  
; APPLICANT: Yutaka KANDA  
; APPLICANT: Mitsuo SATOH  
; APPLICANT: Kazuyasu NAKAMURA  
; APPLICANT: Kazuhisa UCHIDA  
; APPLICANT: Toyohide SHINKAWA  
; APPLICANT: Naoko YAMANE  
; APPLICANT: Motoo YAMASAKI  
; APPLICANT: NO. US20030115614A1uo HANAI  
; TITLE OF INVENTION: ANTIBODY COMPOSITION-PRODUCING CELL  
; FILE REFERENCE: 249-202  
; CURRENT APPLICATION NUMBER: US/09/971,773  
; CURRENT FILING DATE: 2002-08-30  
; PRIOR APPLICATION NUMBER: JP 2000-308526  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 60/268,926  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: Patent in ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 2008  
; TYPE: DNA  
; ORGANISM: Cricetulus griseus  
US-09-971-773-1

Query Match 100.0%; Score 2008; DB 11; Length 2008;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 AACGAAACTTATTTTCTGTGTGCTAAGTACAGAGTACAAATGTTTCCAAATCTT 60  
Db 1 AACGAAACTTATTTTCTGTGTGCTAAGTACAGAGTACAAATGTTTCCAAATCTT 60  
Qy 61 TGAGCTCCGAGACAGACAGAGGAGTTTCAAACTCTGAAATCGCGGCATCGACTGGTTCC 120

Db 61 TGAGCTCCGAGAGGAGGAGTGAAGTCTGAAATCTGAAATCGCGGATGAGCTGTTCC 120  
Qy 121 TGGCGTGGATATGCTCATTTCTTTTCCCTGGGGACCTTATGCTTTATATAGTGCT 180  
Db 121 TGGCGTGGATATGCTCATTTCTTTTCCCTGGGGACCTTATGCTTTATATAGTGCT 180  
Qy 181 CATTTGGTTGAGATTAATGACCACTGACCACTTCTAGCAGAGAACTCTCCAAAGATTCCT 240  
Db 181 CATTTGGTTGAGATTAATGACCACTGACCACTTCTAGCAGAGAACTCTCCAAAGATTCCT 240  
Qy 241 GCAAGCTGGAGCGCTTAAACAAACAAATGAAGACTTGAAGAGAACTGCTGAGTCTCTC 300  
Db 241 GCAAGCTGGAGCGCTTAAACAAACAAATGAAGACTTGAAGAGAACTGCTGAGTCTCTC 300  
Qy 301 CGAATACCAAGAGCGCTTAAAGCAAGAGATTTGAAATTAAGAGAACTGAGTCTGAGTCTC 360  
Db 301 CGAATACCAAGAGCGCTTAAAGCAAGAGATTTGAAATTAAGAGAACTGAGTCTGAGTCTC 360  
Qy 361 GAAACAGCTTTGTAAGGCAAGAGATTTGAAATTAAGAGAACTGAGTCTGAGTCTC 420  
Db 361 GAAACAGCTTTGTAAGGCAAGAGATTTGAAATTAAGAGAACTGAGTCTGAGTCTC 420  
Qy 421 CTGGGAAAGGATCATGAATCTTAAAGGAGGAGATTTGAAATTAAGAGAACTGAGTCTG 480  
Db 421 CTGGGAAAGGATCATGAATCTTAAAGGAGGAGATTTGAAATTAAGAGAACTGAGTCTG 480  
Qy 481 TTTTCTTACAAAGTGAATGAGGAAATTAAGAGAACTGAGTCTGAGTCTC 540  
Db 481 TTTTCTTACAAAGTGAATGAGGAAATTAAGAGAACTGAGTCTGAGTCTC 540  
Qy 541 CATGAGAGTGAATCTTTGAGTATGAGCAATCATGAAAGTCTATCATGACAGATCTA 600  
Db 541 CATGAGAGTGAATCTTTGAGTATGAGCAATCATGAAAGTCTATCATGACAGATCTA 600  
Qy 601 TACTACCTCAGTCAACAGATGAGGAGTGAAGTGGCGGAGAAAGAGCAAGATCTG 660  
Db 601 TACTACCTCAGTCAACAGATGAGGAGTGAAGTGGCGGAGAAAGAGCAAGATCTG 660  
Qy 661 ACAGAGCTGTCACGGGAGATACATATCTGAGAACTCCAGGAGTCCAGGAGTCCAGGAG 720  
Db 661 ACAGAGCTGTCACGGGAGATACATATCTGAGAACTCCAGGAGTCCAGGAGTCCAGGAG 720  
Qy 721 AGAAGCTGCTATGATTAATCAACAAAGCTGTCGATGATGTCACATCATG 780  
Db 721 AGAAGCTGCTATGATTAATCAACAAAGCTGTCGATGATGTCACATCATG 780  
Qy 781 GTTACTGCTTCAATGATGCTTATGAGCACTGAGTCCAGGAGTCCAGGAGTCCAGGAG 840  
Db 781 GTTACTGCTTCAATGATGCTTATGAGCACTGAGTCCAGGAGTCCAGGAGTCCAGGAG 840  
Qy 841 TGGCGCTATGCTGAGGAGTGGAGACTGTTTATGAGCACTGAGTCCAGGAGTCCAGGAG 900  
Db 841 TGGCGCTATGCTGAGGAGTGGAGACTGTTTATGAGCACTGAGTCCAGGAGTCCAGGAG 900  
Qy 901 GACAGCTGCTGCTTCCACTGAGCACTGCTGAGTGAAGTGAAGCAAAATGTTCAA 960  
Db 901 GACAGCTGCTGCTTCCACTGAGCACTGCTGAGTGAAGTGAAGCAAAATGTTCAA 960  
Qy 961 GTGGTGGAGTCCCAATGTAGAGAGCTCCATCTGCTCTTACTTACCTTGGCT 1020  
Db 961 GTGGTGGAGTCCCAATGTAGAGAGCTCCATCTGCTCTTACTTACCTTGGCT 1020  
Qy 1021 GTACAGAGAGCTTGGAGTGGAGACTGTTTATGAGCACTGAGTCCAGGAGTCCAGGAG 1080  
Db 1021 GTACAGAGAGCTTGGAGTGGAGACTGTTTATGAGCACTGAGTCCAGGAGTCCAGGAG 1080  
Qy 1081 GTATCCAGTGTGTCATGATGATGCTGCTGAGTCCAGGAGTCCAGGAGTCCAGGAG 1140  
Db 1081 GTATCCAGTGTGTCATGATGATGCTGCTGAGTCCAGGAGTCCAGGAGTCCAGGAG 1140  
Qy 1141 GAAACCAAGAGCTTGGCTTCAACATCTGAGTCTGAGTCCAGGAGTCCAGGAGCT 1200

Db 1141 GAAACCAAGAGCTTGGCTTCAACATCTGAGTCTGAGTCCAGGAGCT 1200  
Qy 1201 GACAAAGTGGGAAACAGAGAGCTTCCATCCCATTTGAGGAAATACATGGTACAGTTGAA 1260  
Db 1201 GACAAAGTGGGAAACAGAGAGCTTCCATCCCATTTGAGGAAATACATGGTACAGTTGAA 1260  
Qy 1261 GAAATTTTTCAGCTTCTCGAACCGAGAACTGAAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1320  
Db 1261 GAAATTTTTCAGCTTCTCGAACCGAGAACTGAAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1320  
Qy 1321 ACTGATGAGCTTCTTTTAAAGGAGGCAAGAGTACTCAATTAATGAATTTAT 1380  
Db 1321 ACTGATGAGCTTCTTTTAAAGGAGGCAAGAGTACTCAATTAATGAATTTAT 1380  
Qy 1381 AGTGATAACTCTATTTTCTTGGTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1440  
Db 1381 AGTGATAACTCTATTTTCTTGGTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1440  
Qy 1441 CGGGCGGTGATCTCGGATATACATTTCTTCCAGGCTGACTTCTTGTGTGACTTTT 1500  
Db 1441 CGGGCGGTGATCTCGGATATACATTTCTTCCAGGCTGACTTCTTGTGTGACTTTT 1500  
Qy 1501 TCATCCAGGCTCTGAGGTTGCTTATGAATCATGCAACCTGATGATGCTCT 1560  
Db 1501 TCATCCAGGCTCTGAGGTTGCTTATGAATCATGCAACCTGATGATGCTCT 1560  
Qy 1561 GCAAACTTCCATTTCTTGGTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1620  
Db 1561 GCAAACTTCCATTTCTTGGTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1620  
Qy 1621 ATTGAGTCTTATCTTCCAGGCTGAACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1680  
Db 1621 ATTGAGTCTTATCTTCCAGGCTGAACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1680  
Qy 1681 ATCAATGCTGCTGAGGCTGAACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1740  
Db 1681 ATCAATGCTGCTGAGGCTGAACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1740  
Qy 1741 GGAAGAAACAGGCTGATCTTCCATCAAGTCCGAGAGAGATGAGAACTCAAGTCAAGTCAAGTCAAG 1800  
Db 1741 GGAAGAAACAGGCTGATCTTCCATCAAGTCCGAGAGAGATGAGAACTCAAGTCAAGTCAAGTCAAG 1800  
Qy 1801 CCTACATATCTGAGGCTGAACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1860  
Db 1801 CCTACATATCTGAGGCTGAACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1860  
Qy 1861 GTTCAGAGCTTCTGAGGCTGAACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1920  
Db 1861 GTTCAGAGCTTCTGAGGCTGAACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1920  
Qy 1921 GCTCCGAGCTGCTGAGGCTGAACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1980  
Db 1921 GCTCCGAGCTGCTGAGGCTGAACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1980  
Qy 1981 GAAGGCTGCTGCTGAGGCTGAACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2008  
Db 1981 GAAGGCTGCTGCTGAGGCTGAACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2008

## RESULT 2

US-09-971-773-2

; Sequence 2, Application US/09971773

; Publication No. US20030115614A1

; GENERAL INFORMATION:

; APPLICANT: Yutaka KANDA

; APPLICANT: Mitsuo SATOH

; APPLICANT: Kazuyasu NAKAMURA

; APPLICANT: Kazuhisa UCHIDA

; APPLICANT: Toyohide SHINKAWA

; APPLICANT: Naoko YAMANE

; APPLICANT: Motoo YAMASAKI

; APPLICANT: NO. US20030115614A1 HANAI

; TITLE OF INVENTION: ANTIBODY COMPOSITION-PRODUCING CELL



FILE REFERENCE: 249-202  
CURRENT APPLICATION NUMBER: US/09/971,773  
CURRENT FILING DATE: 2002-08-30  
PRIOR APPLICATION NUMBER: JP 2000-308526  
PRIOR FILING DATE: 2000-10-06  
PRIOR APPLICATION NUMBER: US 60/268,926  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 73  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 1728  
TYPE: DNA  
ORGANISM: Mus musculus  
US-09-971-773-2

Query Match 78.9%; Score 1584; DB 11; Length 1728;  
Best Local Similarity 94.8%; Pred. No. 0;  
Matches 1638; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

Qy 100 ATGCGGGCATGACTGGTCTCGGCTGGGATTATGCTCATTTCTTTTGGCTGGGGACC 159  
Db 1 ATGCGGGCATGACTGGTCTCGGCTGGGATTATGCTCATTTCTTTTGGCTGGGGACC 60

Qy 160 TTATTTGTTTATATAGTGGTCAITTTGTTTCGAGATATGACCCCTGACCATTTCTAGC 219  
Db 61 TTGTTATTTTATATAGTGGTCAITTTGTTTCGAGATATGACCCCTGACCATTTCTAGC 120

Qy 220 AGAGAACTCTCCAGATTTCTTGAAAGCTGGAGCGCTTAAACAAACAAATGAAGACTTG 279  
Db 121 AGAGAACTCTCCAGATTTCTTGAAAGCTTGAACGCTTAAACAAACAAATGAAGACTTG 180

Qy 280 AGGAGATGGCTGAGTCTCCGATACCAAGCGCCCTATTGATCGAGGACAGCTACA 339  
Db 181 AGGGAATGGCTGAGTCTCCGATACCAAGCGCCCTATTGATCGAGGACAGCTACA 240

Qy 340 GGAAGAGTCCGTGTTTGAAGAAACAGCTTGTAAAGCCAAAGAACAGATTTGAAAATTAC 399  
Db 241 GGAAGAGTCCGTGTTTGAAGAAACAGCTTGTAAAGCCAAAGAACAGATTTGAAAATTAC 300

Qy 400 AAGAAACAGCTAGGATGATCTGGGAAAGATCATGAAATCTTAAAGAGAGATGAA 459  
Db 301 AAGAAACAGCTAGGATGATCTGGGAAAGATCATGAAATCTTAAAGAGAGATGAA 360

Qy 460 AATGAGCTAAGAGCTCGTGTGTTTCTCAAAAGTGAATTAAGAAATTAAGAAATTA 519  
Db 361 AATGAGCTAAGAGCTCGTGTGTTTCTCAAAAGTGAATTAAGAAATTAAGAAATTA 420

Qy 520 GAAGAAACGAACTCCAAAGACATGCAGATGAAATTTCTTTGGATTTAGGACATCATGAA 579  
Db 421 GAAGAAATGAATCCAAAGACATGCAGATGAAATTTCTTTGGATTTAGGACATCATGAA 480

Qy 580 AGGTCTATGACAGATCTATATCTAAGTCAAGATGAGATGAGGAGGAGTGGCGG 639  
Db 481 AGGTCTATGACAGATCTATATCTAAGTCAAGATGAGATGAGGAGGAGTGGCGT 540

Qy 640 GAAAGAGCCAAAGATCTGACAGAGCTGTCAGCGGAGATTAACATATCTCGAGAT 699  
Db 541 GAAAGAGCCAAAGATCTGACAGAGCTGTCAGCGGAGATTAACATATCTCGAGAT 600

Qy 700 CCCAAGGACTGACGAAAGCAGAAAGCTGTTATGTAATATCAACAAAGGCTGTGGCTAT 759  
Db 601 CCTAAGGACTGACGAAAGCAGAAAGCTGTTATGTAATATCAACAAAGGCTGTGGCTAT 660

Qy 760 GGATGTCAACTCCATCATGTGGTGTATGCTTTCAATGATTTGTCACCCGCAACA 819  
Db 661 GGTGTCAACTCCATCATGTGGTGTATGCTTTCAATGATTTGTCACCCGCAACA 720

Qy 820 CTATCTTGGAAATCTCAGAAATTTGGCGCTATGCTACTGAGGATGGGACATGTGTTTGA 879  
Db 721 CTATCTTGGAAATCTCAGAAATTTGGCGCTATGCTACTGAGGATGGGACATGTGTTTGA 780

Qy 880 CTTGTAAGTGAACATGACAGAGAGTCTGSCCTCTCCACTTGGACATGCTGTGAGGTGAA 939  
Db 1 CTTGTAAGTGAACATGACAGAGAGTCTGSCCTCTCCACTTGGACATGCTGTGAGGTGAA 939

Db 781 CTTGTAAGTGAACATGACAGAGAGTCTGSCCTCTCCACTGAGCACTGTGTAGGTGAA 840

Qy 940 GTGAAGGACAAAAATGTTCAAGTGTGAGCTCCCAATTTGAGACAGCTCCATCCTCGT 999

Db 841 GTAAATGACAAAAACATTCAGTGTGAGCTCCCAATTTGAGACAGCTCCATCCTCGG 900

Qy 1000 CTTCTTTACTTTACCTTTGGCTGTACCAAGAGACCTTCAGATCGACTCTCGAGAGTCCAT 1059

Db 901 CTTCTTTACTTTACCACTGGCTGTTCCAGAGACCTTCAGAGCCGACTCTTAAGAGTCCAT 960

Qy 1060 GGTGATCTCGAGTGTGGTGTATCCAGTTTGTCAAAATCTTGTATCCGTCACAACT 1119

Db 961 GGTGACCTCGAGTGTGGTGTATCCAGTTTGTCAAAATCTTGTATCCGTCACAACT 1020

Qy 1120 TGGCTGGAAGGGAATAGAGAAACCAAGAGCTTGGCTTCAAAATCTTGTATCCGTCACAACT 1179

Db 1021 TGGCTGGAAGGGAATAGAGAAACCAAGAGCTTGGCTTCAAAATCTTGTATCCGTCACAACT 1080

Qy 1180 GGAGTCCATGTTCAGACGACTGACAAAGTGGGAACAGAGCAGCTTCCATCCCATTTGAG 1239

Db 1081 GGAGTCCATGTTCAGACGACTGACAAAGTGGGAACAGAGCAGCTTCCATCCCATTTGAG 1140

Qy 1240 GAATACATGTGTACGTTGAAGAAACATTTTCAGCTTCTCGAACGAGATGAAGTGTAT 1299

Db 1141 GAGTACATGTGTACGTTGAAGAAACATTTTCAGCTTCTCGAACGAGATGAAGTGTAT 1200

Qy 1300 AAAAAAGAGTGTATCTGGCCACTGATGACCTTTCTTTTAAAGGAGGCAAGACAAAG 1359

Db 1201 AAAAAAGAGTGTATCTGGCCACTGATGACCTTTCTTTTAAAGGAGGCAAGACAAAG 1260

Qy 1360 TACTCCAATATGAATTTATTTAGTGATAAATCTTATTTCTTGGTTCAGCTGGACTACAAAC 1419

Db 1261 TACTCCAATATGAATTTATTTAGTGATAAATCTTATTTCTTGGTTCAGCTGGACTACAAAC 1320

Qy 1420 CGATACACAGAAATTCATTTCCGGGGTGTATCTGATATACATTTCTCTCCAGGCT 1479

Db 1321 CGGTACACAGAAATTCATTTCCGGGGTGTATCTGATATACATTTCTCTCCAGGCT 1380

Qy 1480 GACTTCTTGTGTACTTTTTCATCCAGCTGTAGGTTGCTTATGAATCATGCA 1539

Db 1381 GACTTCTTGTGTACTTTTTCATCCAGCTGTAGGTTGCTTATGAATCATGCA 1440

Qy 1540 ACCTGCTATCTGATGCTCTGCAAACTTCATTTCTTTAGATGACATCTACTATTTTGA 1599

Db 1441 ACCCTGCTATCTGATGCTCTGCAAACTTCATTTCTTTAGATGACATCTACTATTTTGA 1500

Qy 1600 GGCAAAATGCCCAACACAGATTTGATCTCTCAACCAACCTCGAACTTAAAGAGGAA 1659

Db 1501 GGCAAAATGCCCAACACAGATTTGATCTCTCAACCAACCTCGAACTTAAAGAGGAA 1560

Qy 1660 ATCCCATGGAACTGAGATATCATTTGGTGTGGCTGGAACCATTTGGAATGTTACTCT 1719

Db 1561 ATTCCCATGGAACTGAGATATCATTTGGTGTGGCTGGAACCATTTGGAATGTTACTCT 1620

Qy 1720 AAAGTGTCAACAGAACTAGGAAAAACAGGCTGTACCTTCTCAAAAGTCCGAGAG 1779

Db 1621 AAAGTGTCAACAGAACTAGGAAAAACAGGCTGTATATCTCTCTACAAAGTCCGAGAG 1680

Qy 1780 AAGATAGAAACAGTCAAAATACCTTACATATCTCTGAAGCTGAAAAATAG 1827

Db 1681 AAGATAGAAACAGTCAAAATACCTTACATATCTCTGAAGCTGAAAAATAG 1728

## RESULT 3

US-10-106-698-381  
; Sequence 381, Application US/10106698  
; Publication No. US20030109690A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide  
; FILE REFERENCE: PA00591  
; CURRENT APPLICATION NUMBER: US/10/106,698  
; CURRENT FILING DATE: 2002-03-27

Qy	1	AACAGAAACTTATTTTCTGTGTGGCTAACTAGAACCCAGAGTACAAATGTTTCCAATTTCTT	60
Db	319	AACAGAAAGTCTATTACCTCTGTCACATAACTAGAAAACAGAGTTTACAAATGTTTTCAAATTTCTT	378
Qy	61	TGAGCTCCGAGAAGACA--GAAGGGAGTTGAAACTCTGAAAATGCGGGCATGGACTGGTT	118
Db	379	TGAGCTCCAGAGCTYACGGGAAGTGAGTTGAAAATCTGAAAATGCGGCCATGGACTGGTT	438
Qy	119	CCTGGCGTTGGATATGATGCTCAATCTTTTGTGCTGGGGACCTTATTTGTTTTTATATAGGTG	178
Db	439	CCTGGCGTTGGAAWTATGCTCAATCTTTTGTGCTGGGGACCTTGCTGTTTTTATATAGGTG	498
Qy	179	GTCAATTTGGTTTCGAGATAATGACCACCTGACCACTTCTAGCAGAGAACTCTCCAAGATTC	238
Db	499	GTCACTTGGTAGAGATATGACCATCTCTGATCACTCTAGCCGAGACTGTCCAAGATTC	558
Qy	239	TTGCAAAAGCTGGAGCGCTTAAAAACAAACAAAATGAAGACTTGAGGAGAAATGGGTG-AGTCT	297
Db	559	TGGCAAAAGCTTGAACGCTTTAAAAACAGCAGAAATGAAGACTTGAGGCGAATGGCCGAAAATCT	618
Qy	298	CTCCGAATACAGAAGGCCCTATTGATCAGGGGACAGCTACAGGAAGAGTCCGTGTTTTTA	357
Db	619	CTCCGGATACAGAAGGCCCTATTGATCAGGGGCCAGCTATAGGAAGAGTAGCGGTGTTTTA	678
Qy	358	GAAGAACAGCTGTGTTAAGGCCAAAGAACAGAGTTGAAAAATTACAAGAAAAACAAGCTAGGAAT	417
Db	679	GAAGAGCAGCTTGTTTAAGGCCCAAGAACAGATTGAAAATTTACAGAAACAGACCAGAAT	738
Qy	418	GATCTGGAAAGGATCATGAAAATCTTAAAGGAGGAGATTGAAAATTGGAGCTAAAGAGCTC	477
Db	739	GCTCTGGGAAAGGATCATGAAAATCTCTGAGGAGGAGATTGAAAATTGGAGCTAAAGAGCTC	798
Qy	478	TGGTTTTTTCTACAAGTGAATTGAAGAAATTTAAGAAATTTAGAGGAAACGAATCTCCAA	537
Db	799	TGGTTTTTTCTCAAGAGTGAATTTGAAGAAATTTAAGAAATTTAGAGGAAATGAATCTCCAA	858
Qy	538	AGACATGCAGATGAAATTTCTTTTGGATTTAGGACATCATGAAAGGTCTTATCATGACAGAT	597
Db	859	AGACATGCAGATGAAATTTCTTTTGGATTTAGGACATCATGAAAGGTCTTATTAATGACGAT	918
Qy	598	CTATATCTACCTCAGTCAAAACAGATGGACAGAGTGATGGCGGGAATAAAGAGCCAAAGAT	657
Db	919	CTATATCTACCTCAGTCAAGACAGATGGACAGAGTGATGGCGGGAATAAAGAGCCAAAGAT	978
Qy	658	CTGACAGAGCTGGTCCAGCGCAGAAATAACATATCTGCAGAAATCCCACGAGCTGCAGCAAA	717

Db 2059 TTGGGAAGCGGCGCTATATCCCTCTCTCAAAAGTTTCAGAGAAAGATAGAAACGGTCAAG 2118  
Qy 1798 TACCTACATATCTGAAGCTGAAATATAGATGAGTGTAAAGAGATTAA----- 1848  
Db 2119 TACCCACATATCTGAGCGCTGAGAAATAAAGCTCAGATGGAAGAGATAAAGCACAAC 2178  
Qy 1849 -----CAACAGAAATTTAGTTTCAGACCATCTCAGCCAGCAGAGAACCCAG----- 1893  
Db 2179 TCAGTTCCAGCAACTCAGTTTCAAAACCATTTCCAGCCAACTGTAGTGAAGAGGCTCTG 2238  
Qy 1894 -----ACTAACATATGTTTCATTTGACAGACATGCTCCGCACCAAGAGCAAGTGGGAACCC 1948  
Db 2239 ATCTAACAAAATAAGGTTATATGAGTAGATCTCTCAGCACCAGCAGCTGGGAACTG 2298  
Qy 1949 TCAGATGCTGC-ACTGGTGGAAAGCGCTCTTTGTGAAGGCGTCTGCTGCCCTCAAGCCCAT 2007  
Db 2299 ACATAGGCTTCAATTTGGTGGAAATTCCTCTTTTAAACAAAGGCGTGAATGCCCTCATACCAT 2358  
Qy 2008 G 2008  
Db 2359 G 2359

## RESULT 4

US-09-839-136-9  
; Sequence 9, Application US/09839136  
; Patent No. US20020081694A1  
; GENERAL INFORMATION:  
; APPLICANT: Naoyuki TANIGUCHI et al.  
; TITLE OF INVENTION: ALPHA 1-6 FUCOSYLTRANSFERASE  
; FILE REFERENCE: 2356-7  
; CURRENT APPLICATION NUMBER: US/09/839,136  
; CURRENT FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: 09/442,629  
; PRIOR FILING DATE: 1999-11-18  
; PRIOR APPLICATION NUMBER: 08/913,805  
; PRIOR FILING DATE: 1998-01-07  
; PRIOR APPLICATION NUMBER: PCT/JP97/00171  
; PRIOR FILING DATE: 1997-01-23  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 2100  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (198)...(1925)  
US-09-839-136-9

Query Match 78.2%; Score 1569.6; DB 9; Length 2100;  
Best Local Similarity 90.1%; Pred. No. 0;  
Matches 1692; Conservative 0; Mismatches 184; Indels 2; Gaps 1;  
Qy 1 AACGAAACTATTTTCTGTGTGCTTAACAGAACAGAGTACAAATGTTTCCAAATCTT 60  
Db 97 AACGAAAGTCTATTCACCTGTGCATCACTAGAAACAGAGTTACAAATGTTTCAATCTT 156  
Qy 61 TGAGCTCCGAGAGACA--GAAGGAGTTGAAACTCTGAAATCGGGCATGAGCTGGTT 118  
Db 157 TGAGCTCCAGGACTCCAGGGAAGTGAAGTTGAAATCTGAAATGCGGCATGAGCTGGTT 216  
Qy 119 CTGGCGTTGGATTATGCTCATTCTTTTGGCTGGGGACCTTATTTTATATAGTG 178  
Db 217 CTGGCGTTGGATTATGCTCATTCTTTTGGCTGGGGACCTTGTCTGTTTATATAGTG 276  
Qy 179 GTCAATTTGTTCCGATATGACCACTGACCACTTCTAGAGAGAACTCTCCAGATTTC 238  
Db 277 GTCACTTGTGATGATATGACCACTTCTGATCCTAGCGGAACTGTCAGAGATTC 336  
Qy 239 TTGCAAGCTGAGCGCTTAAACCAACAAATGAAACACTTGAAGGAGATGGCTGAGTCTC 298  
Db 337 TGGCAAGCTTGAACTTAAACAGCAGATGAGACTTGAAGGCAATGCGCAATCTC 396

Qy 299 TCCGAATACCAAGAGGCCCTATTGATCAGGGGACAGCTACAGGAAGAGTCCGTTGTTTAG 358  
Db 397 TCCGGATACCAAGAGGCCCTATTGATCAGGGGCCAGCTATAGGAAGAGTAGTACGCGTTTAG 456  
Qy 359 AAGAACAGCTTGTAAAGGCCAAAGAACAGATTGAAATTTACAGAAACAGAGCTAGGATG 418  
Db 457 AAGAGCAGCTTGTAAAGGCCAAAGAACAGATTGAAATTTACAGAAACAGAGCTAGGATG 516  
Qy 419 ATCTGGGAAGGATCATGAAATCTTAAAGGAGGAGATTGAAATTTGAGAGCTCT 478  
Db 517 GTCTGGGAAGGATCATGAAATCTTGAAGGAGGAGATTGAAATTTGAGAGCTCT 576  
Qy 479 GGTTTTTTCTACAAAGTGAATTTGAAGAAATTTAAAGAAATTTAGAAAGGAAACGAACTCCAAA 538  
Db 577 GGTTTTTTCTACAGATGAATTTGAAGAAATTTAAAGAACTTTAGAAAGGAAATGAACCTCCAAA 636  
Qy 539 GACATGAGATGAATTTCTTTTGGATTTAGACATCATGAAAGTCTATCATGACAGATC 598  
Db 637 GACATGAGATGAATTTCTTTTGGATTTAGGACATCATGAAAGTCTATATATGACGGATC 696  
Qy 599 TATACTACCTCAGTCARACAGATGAGCAGCTGAGTGGCGGAAAAAGAACCCCAAAGATC 658  
Db 697 TATACTACCTCAGTCAGACAGATGAGCAGCTGAGTGGCGGAAAAAGAGGCCAAAGATC 756  
Qy 659 TGACAGAGCTGGTCCAGCGGAGAAATAACATATCTGCAGAAATCCCAAGGACTGCGAGCAAG 718  
Db 757 TGACAGAACTGGTTCAGCGGAGATACATATCTTCAGAACTCCCAAGGACTGCGAGCAAG 816  
Qy 719 CCAGAAAGCTGGTATGTAATATCAAAAGGCTGTGGCTATGGATGTCAATCCATCATG 778  
Db 817 CCAAAAGCTGGTGTGTAATATCAAAAGGCTGTGGCTATGGCTGTGAGCTTCCATCATG 876  
Qy 779 TGGTTTACTGCTTCATGATTTGCTTATGGCAAGGAGTGGTGTAGACCTGTAGTGAATCTCAGA 838  
Db 877 TGGTCTACTGCTTCATGATTTGCTTATGGCAAGGAGTGGTGTAGTGAATCTCAGA 936  
Qy 839 ATTGGCGCTATGCTACTGGAGGATGGGAGACTGTGTTTAGACCTGTAGTGAATGAGATGCA 898  
Db 937 ATTGGCGCTATGCTACTGGTGGATGGGAGACTGTATTTAGGCGCTGTAGTGAATGAGATGCA 996  
Qy 899 CAGACAGCTTGGCTCTCCACTGACACTGGTCAAGTGAAGTGAAGGAGCAAAAATGTTTC 958  
Db 997 CAGACAGATCTGGCATCTCCACTGACACTGGTCAAGTGAAGTGAAGGAGCAAAAATGTTTC 1056  
Qy 959 AAGTGGTGGAGTCCCATTTGTAGACAGCTCCATCTCGTCTCTCTTACTTACCTTGG 1018  
Db 1057 AAGTGGTGGAGTCCCATTTGTAGACAGCTCTTCATCCCGCTCTCCATATTTACCTTGG 1116  
Qy 1019 CTGTACCAAGAGAGCTTGCAGATCGACTCTTCAGAGTCCATGTTGATCCTCGAGTGTGGT 1078  
Db 1117 CTGTACCAAGAGAGCTTGCAGATCGACTTGTACAGTGCATGGTGACCTCGAGTGTGGT 1176  
Qy 1079 GGGTATCCAGTGTGTCAAATACTTGTATCCGTCCACAACTTGGCTGGAAGGAGAAATAG 1138  
Db 1177 GGGTGTCTCAGTGTGTCAAATACTTGTATCCGCCACAGCTTGGCTAGAAAAGAAATAG 1236  
Qy 1139 AAGAAACCAAGAGAGAGTGGCTTCAAACATCAGTTATTTGGAGTCCATGTCAGACGCA 1198  
Db 1237 AAGAAAGCAGCAAGAGAGTGGCTTCAAACATCAGTTATTTGGAGTCCATGTCAGACGCA 1296  
Qy 1199 CTGACAAAGTGGGAACAGAGCAGCTTCCATCCATTTGAGGAAATACATGTTGATGATG 1258  
Db 1297 CAGCAAAAGTGGGAACAGAGCTGCTTCCATCCATTTGAAGAGTACATGGTGCATGTTG 1356  
Qy 1259 AAGAACTATTTTTCAGCTTCTCGAAGCGCAAGTGAAGTGAATGAAAGAGTGTATCTGG 1318  
Db 1357 AAGAACTATTTTTCAGCTTCTCGAAGCGCAAGTGAAGTGAATGAAAGAGTGTATTTGG 1416  
Qy 1319 CCATGATGACCTTCTTTTGAAGGAGGCAAGACAAAGTACTTCCAAATTTATGAAATTTA 1378  
Db 1417 CCAGATGACCTTCTTTTGAAGGAGGCAAGACAAAGTACTTCCAAATTTATGAAATTTA 1476

1379	Qy	TTAGTGTAACTCTATTTCTTTGGTCAGCTGGACTACACAACCGGATACACAGAAAATTCAC	1438
1477	Db	TTAGTGATAACTCTATTTCTTGSTGCTGCAATTCGATACACAGAAAATTCAC	1536
1439	Qy	TTCCGGGGGNGATCCCTGGGATATACATTTCTCCCAAGGCTGACTTCCTTGTTGTACTT	1498
1537	Db	TTCCGTGGAGTGATCCCTGGATATACATTTCTCTCCAGGCAGACTTCCTAGTGTGTACTT	1596
1499	Qy	TTTTCATCCCAAGGTCTGTGAGGGTTGCTTTATGAATCATGCACAACTGCATCTCTGATGCCT	1558
1597	Db	TTTTCATCCCAAGGTCTGTGAGGTTACTTTATGAAATTTAGCAAACTACATCTCTGATGCCT	1656
1559	Qy	CTGCAAACTTCCATCTTTTAGATGACATCTACTATTTTGGAGGCCAAAATGCCCAACAAC	1618
1657	Db	CTGCAAACTTCCATCTTTTAGATGACATCTACTATTTTGGGGGCCAGAAATGCCCACAATC	1716
1619	Qy	AGATTGCAGTTTATCTTCACAACTTCGAACCTAAAGAGGAAATCCCATCGAACCTGGAG	1678
1717	Db	AAATTGCCATTTATGTCTCACCACCCCGAACTGCAGATGAAATTTCCCATGGAACCTGGAG	1776
1679	Qy	ATATCATTTGGTGTGGCTGGAAACCAATTGGAATGGTTACTCTTAAAGTGTCACACAGAAAAC	1738
1777	Db	ATATCATTTGGTGTGGCTGGAAATCAATTGGGATGGCTATTCTTAAAGGTGTCAAACAGGAAAT	1836
1739	Qy	TAGAAAAACAGGCCTGTACCCCTTCTACAAAGTCCGAGAGAGATAGAAAACAGTCAAAT	1798
1837	Db	TGGGAGGACGGGCTATATCTCTCTACAAAGTTCAGAGAGAGATAGAAAACGCTCAAGT	1896
1799	Qy	ACCTCATATATCTGGAAGCTGAAAAATAGAGATGGAGTGTAGAGATTAAACAAGAAATT	1858
1897	Db	ACCCACATATCTGTAGGCTGAGAAATAAAGCTCACATGGAAGAGATAAACCGCAAACT	1956
1859	Qy	TAGTTTCAGACCATCTCAG	1876
1957	Db	CAGTTTCGACCAAACTCAG	1974

## RESULT 5

```

US-09-839-136-1
; Sequence 1, Application US/09839136
; Patent No. US20020081694A1
; GENERAL INFORMATION:
; APPLICANT: Naoyuki TANIGUCHI et al.
; TITLE OF INVENTION: ALPHA 1-6 FUCOSYLTRANSFERASE
; FILE REFERENCE: 2356-7
; CURRENT APPLICATION NUMBER: US/09/839,136
; CURRENT FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 09/442,629
; PRIOR FILING DATE: 1999-11-18
; PRIOR APPLICATION NUMBER: 08/913,805
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: PCT/JP97/00171
; PRIOR FILING DATE: 1997-01-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1728
; TYPE: DNA
; ORGANISM: Pig
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1728)
US-09-839-136-1

```

Db	61	TTGCTATTTTACATAGGTGGTCACTTGGTACGAGATAATGACCACTCTGATCACTCTAGC	120
Qy	220	AGAGAACTCTCCAAAGATTTCTTGCAAAAGCTGGAGCGCTTAAAAACAACAAATGAAGACTTGG	279
Db	121	CGAGAACTGCTCCAAAGATTTTGGCAAAAGCTGGAAAGCTTAAAAACAACAAATGAAGACTTGG	180
Qy	280	AGGAGAAATGGCTGAGTCTCTCCGAATACCAAGAAAGGCCCTATTGTATCAGGGGACAGCTACA	339
Db	181	AGGAGAAATGGCTGGATCTCTCCGAATACCAAGAAAGGCCCAATTGTATCAGGGGCGAGCTTCA	240
Qy	340	GGAAAGTCCGTGTTTTAGNAGAACAGCTGTGTTAAGGCCAAAGAAACAGATTGAAATATAC	399
Db	241	GGAAAGTTCGTGCTTTAGAAAGCAATTTATGAAGGCCAAAGAAACAGATTGAAATATAT	300
Qy	400	AAGAAACAAGCTTAGGAATGATCTGGCAAAAGGATCATGAATCTTAAGGAGGAGGATTGAA	459
Db	301	AAGAAACAACACTAAAAAATGGTCCAGGGAAGGATCATGAATCTTAAGGAGGAGGATTGAA	360
Qy	460	AATGGAGCTAAAGAGCTCTGGTTTTTCTACAAAGTGAAATGAAGAAATTAAGAAATTA	519
Db	361	AATGGAGCTAAAGAGCTCTGGTTTTTCTACAAAGTGAGTTGAAGAAATTAAGAAATTTA	420
Qy	520	GAAGGAAAAGAACTCCAAAGACATGCAGATGAATTTCTTTTGGATTTAGACATCATGAA	579
Db	421	GAAGGAAATGAATCTCCAAAGACATGCAGATGAATTTCTTATCAGATTTGGGACATCATGAA	480
Qy	580	AGGTCTATCATGACAGATCTATACTACCTCAGTCAAACAGATGGACGAGCTAGTGGCGG	639
Db	481	AGGTCTATCATGACAGATCTATACTACCTCAGTCAAACAGATGGGCGAGCTGATTTGGCGT	540
Qy	640	GAAGGAAAGCCAAAGATCTGAACAGAGCTGGTCCAGCGGAGAAATAATATCTGCAGAAAT	699
Db	541	GAAGGAGGCCAAAGATCTGAACAGAGCTGGTCCAGCGGAGAAATAATATCTGCAGAAAT	600
Qy	700	CCCAAGGACTGCAGCAAAAGCCAGAAAGCTGGTATGTAATATCAACAAAGGCTGTGGCTAT	759
Db	601	CCCAAGGACTGCAGCAAAAGCCAAAGAGCTAGTGTGTAATCAACAAAGGCTGTGGCTAT	660
Qy	760	GGATGCAACTCCATCATGTGGTTTATCTGCTTCATGATTGCTTATGGCCACCCAGCGAAACA	819
Db	661	GGCTGTGAGCTCCATCATGTAGTGTATCTGCTTTATGATTTGCATATGCGACCCAGCGAAACA	720
Qy	820	CTCATCTTGGAAATCTCAGAAATGGCGCTATGCTACTGGAGGATGGGAGACTGTGTTTAGA	879
Db	721	CTCGCCTTGGAAATCTCACAATGGCGCTACGCTACTGGGGGATGGGAAACTGTGTTTAGA	780
Qy	880	CTGTATGATGAGACATGCAACAGCTGTGGCTCTCCACTGGACACTGGTCAGGTGAA	939
Db	781	CCTGTATGATGAGACGTGCAACAGATCTGGCGAGCTCCACTGAGCAATTTGGTCAGGTGAA	840
Qy	940	GTGAGGACAAAATGTTCAAGTGGTGCAGCTCCCATTTGTAGACAGCTCCATCTCGT	999
Db	841	GTAAGGACAAAATGTTCAAGTGGTGTAGCTCCCATTTGTAGACAGTGTTCATCTCGT	900
Qy	1000	CCTCCTTACTTACCTTTGGCTGTACCAGAAAGCTTTGCAGATCGACTCTCTGAGAGTCCAT	1059
Db	901	CCTCCATATTTACCCCTGGCTGTCCCAAGAGCTTTGAGATCGACTTGTACAGGTCCAT	960
Qy	1060	GGTGATCCTGCAGTGTGGGGTATCCAGTTTGTCAAAATACATTGTATCCGTCCCAACCT	1119
Db	961	GGTGATCCTGCAGTGTGGGGTATCCAGTTTGTCAAAATACATTGTATCCGTCCCAACCC	1020
Qy	1120	TGGCTGGAAGGGAATAGNAGAAACCAACAAGCTTTGGCTTTCAAACATCCAGTTATT	1179
Db	1021	TGGCTGGAAGGGAATAGNAGAGGGCCACCAAGAGCTAGGCTTTCAAACATCCAGTTATT	1080
Qy	1180	GGAGTCCATGTCAGACGACTGACAAAGTGGGAAACAGAAAGCAGCTTCCATCCCATTTGAG	1239
Db	1081	GGAGTCCATGTTAGACGACAGACAAAGTGGGAGCGNAGCAGCTTCCATCCCATTTGAG	1140
Qy	1240	GAATACATGTTACAGTTTGAAGAAATTTTCAGCTTCTCGAACGCAAGATGAAAGTGGAT	1299

Db 1141 GAATACACGGTGACGTTGAAGAAGACTTTTCAGTCTTCTGCTCGAGAAATGCAAGTGGAT 1200  
Qy 1300 AAAAAAAGAGTGTATCTGCGCACTGATGACCCCTCTCTTTGTTAAAGGAGCGCAAGCAAG 1359  
Db 1201 AAAAAAAGGGTGTATTTGGCCACAGATGACCTCTGTTTGTAAAGAGGCGCAAAACAAG 1260  
Qy 1360 TACTCCAATATGAATTTATTAGTGATAACTCTATTCTTTGGTGCAGCTGGACTACACAAC 1419  
Db 1261 TACCCAGTATGAATTTATTAGTGATAACTCTATTCTTTGGTGCAGCTGGACTACATAAT 1320  
Qy 1420 CGATACACAGAAATTCATCTCGGGCGTGATCCTGGATATACACTTTCTCTCCAGGCT 1479  
Db 1321 CGATATACAGAAATTCATCTCGGGGTGTGATCTCGGATATACACTTTCTCTCCAGGCA 1380  
Qy 1480 GACTTCTCTGTGTGTAATTTTCATCCAGGCTGTAGGGTGTGATGAATAATCATGCAA 1539  
Db 1381 GACTTCTCTAGTGTGTAATTTTCATCGAGGCTGTAGGTTGCTTATGAATAATCATGCAA 1440  
Qy 1540 ACATGCTATCTGATCGCTCTGCAAACTTCCATCTTTTAGATGACATCTACTATTTTGA 1599  
Db 1441 GCGCTGCATCTGATGCTCTGCGAACTTCCGTTCTTTGGATGACATCTACTATTTTGA 1500  
Qy 1600 GGCACAAATGCCACAAACAGATTCCAGTTTATCTCTACCAACCTCGAACTAAAGAGAA 1659  
Db 1501 GGCACAAATGCCACAAACAAATGCCATTTATCTCTACCAACCTCGAACTGAAGGAGAA 1560  
Qy 1660 ATCCCATGGAACTGGAGATATCATTTGTTGGCTGGAAACCATTTGGAATGGTTACTCT 1719  
Db 1561 ATCCCATGGAACTGGAGATATTAATTTGTTGGCTGGAAATCATCTGGATGGCTATCCT 1620  
Qy 1720 AAGGTGTCAACAGAAAACTAGAAAAACAGGCTGTACCCCTTCTCAAAAAGTCCGAGAG 1779  
Db 1621 AAGGTGTTAACAGAAAACTGGAGGAGCGGCTTATCTCTCTCAAAAAGTTCGAGAG 1680  
Qy 1780 AAGTAGAAACAGTCAATATACCTACATATCTCTGAAGCTGAAAAATA 1826  
Db 1681 AAGTAGAAACAGTCAAGTACCCACATATCCCGAGGCTGCAAGTA 1727

RESULT 6

US-09-971-773-6  
; Sequence 6, Application US/09971773  
; Publication No. US20030115614A1  
; GENERAL INFORMATION:  
; APPLICANT: Yutaka KANDA  
; APPLICANT: Mitsuo SATOH  
; APPLICANT: Kazuyasu NAKAMURA  
; APPLICANT: Kazuhisa UCHIDA  
; APPLICANT: Toyohide SHINKAWA  
; APPLICANT: Naoko YAMANE  
; APPLICANT: Motoo YAMASAKI  
; APPLICANT: No. US20030115614A1uo HANAI  
; TITLE OF INVENTION: ANTIBODY COMPOSITION-PRODUCING CELL  
; FILE REFERENCE: 249-202  
; CURRENT APPLICATION NUMBER: US/09/971,773  
; PRIORITY FILING DATE: 2002-08-30  
; PRIOR APPLICATION NUMBER: JP 2000-308526  
; PRIORITY FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 60/268,926  
; PRIORITY FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 979  
; TYPE: DNA  
; ORGANISM: Cricetulus griseus  
US-09-971-773-6

Query Match 48.6%; Score 976.4; DB 11; Length 979;  
Best Local Similarity 99.9%; Pred.No. 3.4e-273;  
Matches 977; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 819 ACTCATCTTGAATCTCAGAATTGGCGCTATGCTACTGGAGGATGGGAGACTGTGTTTAG 878

Db 1 ACTCATCTTGAATCTCAGAATTGGCGCTATGCTACTGGAGGATGGGAGACTGTGTTTAG 60  
Qy 879 ACCTGTAAGTCAGACATGSCACAGACAGGTCTGGCTCTCCACTGACACTGCTCAGGTGA 938  
Db 61 ACCTGTAAGTCAGACATGSCACAGACAGGTCTGGCTCTCCACTGACACTGCTCAGGTGA 120  
Qy 939 AGTGAAGACAAAAAATGTTCAAGTGGTCCAGCTCCCAATTGTTAGACAGCTCCATCTCG 998  
Db 121 AGTGAAGACAAAAAATGTTCAAGTGGTCCAGCTCCCAATTGTTAGACAGCTCCATCTCG 180  
Qy 999 TCCTCTTACTTACCTTGGCTGTACACAGAAAGACCTTGCAGATCGACTCTCTGAGAGTCA 1058  
Db 181 TCCTCTTACTTACCTTGGCTGTACACAGAAAGACCTTGCAGATCGACTCTCTGAGAGTCA 240  
Qy 1059 TGGTATCTCTCGAGTGGTGGGTATCCAGTGTGTCAAAATGTTGATCCGTCACAAC 1118  
Db 241 TGGTATCTCTCGAGTGGTGGGTATCCAGTGTGTCAAAATGTTGATCCGTCACAAC 300  
Qy 1119 TTGGCTGAAAAGGAAATAGAAAGAAACCAACAAAGAGCTTGGCTTCAAAACATCCAGTTAT 1178  
Db 301 TTGGCTGAAAAGGAAATAGAAAGAAACCAACAAAGAGCTTGGCTTCAAAACATCCAGTTAT 360  
Qy 1179 TGGAGTCATGTCAGACGCACTGACAAAGTGGGAAACAGAGAGGCTTCCATCCCATGA 1238  
Db 361 TGGAGTCATGTCAGACGCACTGACAAAGTGGGAAACAGAGAGGCTTCCATCCCATGA 420  
Qy 1239 GGAATACATGTAACAGTTGAAGAACATTTTCAGCTTCCGAAACGCAAGAAAGTGA 1298  
Db 421 GGAATACATGTAACAGTTGAAGAACATTTTCAGCTTCCGAAACGCAAGAAAGTGA 480  
Qy 1299 TAAAAAAGAGTGTATCTGGCCACTGATGACCTCTTTGTTTAAAGAGGCAAAAGACAAA 1358  
Db 481 TAAAAAAGAGTGTATCTGGCCACTGATGACCTCTTTGTTTAAAGAGGCAAAAGACAAA 540  
Qy 1359 GTACTCCAATATGAATTTATTAGTGATACTCTATTTCTTGGTCAGCTGAGTACACAA 1418  
Db 541 GTACTCCAATATGAATTTATTAGTGATACTCTATTTCTTGGTCAGCTGAGTACACAA 600  
Qy 1419 CCGATACACAAAAATTCACCTCGGGCGTGATCTCGATATACACTTTCTCTCCAGGC 1478  
Db 601 CCGATACACAAAAATTCACCTCGGGCGTGATCTCGATATACACTTTCTCTCCAGGC 660  
Qy 1479 TGACTTCTCTGTGTACTTTTTCATCCAGGTCTGTAGGGTGTGCTTATGAATCATGCA 1538  
Db 661 TGACTTCTCTGTGTACTTTTTCATCCAGGTCTGTAGGGTGTGCTTATGAATCATGCA 720  
Qy 1539 AACACTGCATCTGATGCTCTCGAAACTTCCATTTTGTAGATGACATCTACTATTTGG 1598  
Db 721 AACACTGCATCTGATGCTCTCGAAACTTCCATTTTGTAGATGACATCTACTATTTGG 780  
Qy 1599 AGGCAAAATGCCACAAACAGATGCGAGTTTATCTCCACCAACCTCGAACTAAAGAGA 1658  
Db 781 AGGCAAAATGCCACAAACAGATGCGAGTTTATCTCCACCAACCTCGAACTAAAGAGA 840  
Qy 1659 AATCCCATGAACTCGAGATATCATTTGGTGTGCTGGAACCATTTGGAATGTTACTC 1718  
Db 841 AATCCCATGAACTCGAGATATCATTTGGTGTGCTGGAACCATTTGGAATGTTACTC 900  
Qy 1719 TAAAGGTGTCACAGAAACTAGGAAAAACAGGCTGTACCTTCTCAAAAGTCCGAGA 1778  
Db 901 TAAAGGTGTCACAGAAACTAGGAAAAACAGGCTGTACCTTCTCAAAAGTCCGAGA 960  
Qy 1779 GAAGATGAACACAGTCAA 1796  
Db 961 GAAGATGAACACAGTCAA 978

RESULT 7  
US-09-971-773-7  
; Sequence 7, Application US/09971773  
; Publication No. US20030115614A1  
; GENERAL INFORMATION:

```
/ APPLICANT: Yutaka KANDA
/ APPLICANT: Mitsuo SATOH
/ APPLICANT: Kazuyasu NAKAMURA
/ APPLICANT: Kazuhisa UCHIDA
/ APPLICANT: Toyohide SHINKAWA
/ APPLICANT: Naoko YAMANE
/ APPLICANT: Motoo YAMASAKI
/ APPLICANT: No. US20030115614ALUO HANAI
/ TITLE OF INVENTION: ANTIBODY COMPOSITION-PRODUCING CELL
/ FILE REFERENCE: 249-202
/ CURRENT APPLICATION NUMBER: US/09/971,773
/ CURRENT FILING DATE: 2002-08-30
/ PRIOR APPLICATION NUMBER: JP 2000-308526
/ PRIOR FILING DATE: 2000-10-06
/ PRIOR APPLICATION NUMBER: US 60/268,926
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 73
/ SOFTWARE: Patentin Ver. 2.1
/ SEQ ID NO 7
/ LENGTH: 979
/ TYPE: DNA
/ ORGANISM: Rattus norvegicus
US-09-971-773-7

Query Match      44.5%; Score 893.2; DB 11; Length 979;
Best Local Similarity 94.6%; Pred. No. 5.8e-249;
Matches 925; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 819 ACTCATCTTGGAAATCTCAGAAATGGCGCTATGCTACTGAGGATGGGAGACTGTTTAG 878
DB 1 ACTCATCTTGGAAATCTCAGAAATGGCGCTATGCTACTGAGGATGGGAGACTGTTTAG 60
QY 879 ACCTGTAAGTAGAGATGCACAGACAGAGTCTGGCTCTCCACTGCACACTGTCAGTGA 938
DB 61 ACCTGTAAGTAGAGATGCACAGACAGATCTGGCTCTCCACTGCACACTGTCAGTGA 120
QY 939 AGTGAAGGACAAAATGTTTCAAGTGGTGGAGTCCCGCTATGTCAGACGCTCCATCCTCG 998
DB 121 AGTGAATGACAAAATATTTCAAGTGGTGGAGTCCCGCTATGTCAGACGCTCCATCCTCG 180
QY 999 TCCTCTTACTTACCTGGCTGTACGAGAGACCTTCGAGATCGACTCTCGAGAGTCCA 1058
DB 181 GCCTCTTACTTACCACTGGCTGTTCCAGAGACCTTCGAGATCGACTCGTAAGAGTCCA 240
QY 1059 TGGTGCATCTCGCAGTGGTGGGTATCCAGTTCGTCAAATCTTGATCCGTCACCAACC 1118
DB 241 TGGTGCATCTCGCAGTGGTGGGTATCCAGTTCGTCAAATCTTGATCCGTCACCAACC 300
QY 1119 TTGGCTGGAAGGGAAATAGAAAGAACCAAGAGCTTGGCTTCAAACATCCAGTTAT 1178
DB 301 TTGGCTAGAAAAGGAAATAGAAAGAGCCACCAAGAGCTTGGCTTCAAACATCCAGTCA 360
QY 1179 TGGAGTCCATGTACAGCGCATGTACAAAGTGGGAACAGAGAGCTTCCATCCCATGA 1238
DB 361 TGGAGTCCATGTACAGCGCATGTACAAAGTGGGAACAGAGAGCTTCCATCCCATGA 420
QY 1239 GGAATACATGTPACAGTGTGAAGAACATTTTCAGCTTCTCGAAAGCAGAAATGAAGTGA 1298
DB 421 AGAGTACATGGTACATGTTTGAAGAACATTTTCAGCTTCTCGAAAGCAGAAATGAAGTGA 480
QY 1299 TAAAAAAGAGGTATCTGGCCACTGATGACCCCTTTTGGTTTAAAGAGGGCAAGACAAA 1358
DB 481 TAAAAAAGAGGTATCTGGCTACCGATGACCCCTTTTGGTTTAAAGAGGGCAAGACAAA 540
QY 1359 GTACTCCAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1418
DB 541 GTACTCCAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 600
QY 1419 CGGTACACAGAAAATTCACCTCGGGGGTGTATCTGGATATACACTTTCTCTCCAGGC 1478
DB 601 TCGGTACACAGAAAATTCACCTCGGGGGTGTATCTGGATATACACTTTCTCTCCAGGC 660
QY 1479 TGACTTCCTTGTGTGCTTTTTCATCCAGGCTGTGTAGGGTGTCTTATGAATCATGCA 1538
```

```
DB 661 TGACTTCCTAGTGTGTAATTTTTCATCCAGGCTGTGCGGTTGCTTTATGAATCATGCA 720
QY 1539 AACACTGCATCCTGATGCTCTGCAAACTTCCTATCTTTTAGATGACATCTACTATTTGG 1598
DB 721 AACCTGATCCTGATGCTCTGCAAACTTCCTATCTTTTAGATGACATCTACTATTTGG 780
QY 1599 AGGCCAAAATGCCCAACACAGATTGCGTTTATCTCCACCACTCGAACTGATGAGGA 1658
DB 781 AGGCCAAAATGCCCAACACAGATTGCGTTTATCTCCACCACTCGAACTGATGAGGA 840
QY 1659 AATCCCATGGAACCTGGAGATATCATTTGTTGGCTGGAAACCTTGGATGTTTACTC 1718
DB 841 AATTCGAATGGAACCTGGAGATATCATTTGTTGGCTGGAAACCTTGGATGTTTACTC 900
QY 1719 TAAAGGTGTCAACAGAAAACCTAGGAAAACAGGCTGTATATCCCTCTACAAAGTCCGAGA 1778
DB 901 TAAAGGTGTCAACAGAAAACCTTGGAAAACAGGCTGTATATCCCTCTACAAAGTCCGAGA 960
QY 1779 GAAGATAGAAACACTCAA 1796
DB 961 GAAGATAGAAACGCTCAA 978

RESULT 8
US-09-918-995-33468
; Sequence 33468, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 33468
; LENGTH: 422
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-33468
```

```
Query Match      18.3%; Score 368.2; DB 11; Length 422;
Best Local Similarity 92.2%; Pred. No. 2.5e-96;
Matches 388; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 863 GGGAGACTGTGTTTAGACCTGTAAAGTGAGACATGCACAGACGCTGTGGCTCTCCACTG 922
DB 2 GGGAGACTGTATTTAGGCTGTAAAGTGAGACATGCACAGACGATCTGGCATCTCCACTG 61
QY 923 GACACTGTCAGTGAAGTGAGGACAAAATGTTCAAGTGGTTCAGTCTCCGCTCCCATTTAG 982
DB 62 GACACTGTCAGTGAAGTGAGGACAAAATGTTCAAGTGGTTCAGTCTCCGCTCCCATTTAG 121
QY 983 ACAGCTCCATCTCTGCTCTTACTTTACCTTTGGCTGTACCAAGAGACCTTGCAGATC 1042
DB 122 ACAGTCTTATCCCGCTCTCTCCATATTTACCTTTGGCTGTACCAAGAGACCTTGCAGATC 181
QY 1043 GACTCTCTGAGAGTCCATGGTGATCCTGAGTGTGGTATCCAGTTTCTCAAAATCT 1102
DB 182 GACTTGTACGAGTGCATGGTGACCTGTCAGTGTGGTGTCTCAGTTTCTCAAAATCT 241
QY 1103 TGATCCGTCACAACTTGGCTGGAAGGGAATAGAGAAACCAAGAGCTTGGCT 1162
DB 242 TGATCCGTCACAACTTGGCTGGAAGGGAATAGAGAAACCAAGAGCTTGGCT 301
QY 1163 TCAAACTCAGTATTTGGAGTCCATGTTCAGACGCTGTACAAAAGTGGAAACAGAGCAG 1222
DB 302 TCAAACTCAGTATTTGGAGTCCATGTTCAGACGCTGTACAAAAGTGGAAACAGAGCAG 361
```



Qy	1223	CGTTCATCCCATGACGAATACATGTCACGTTGAAGAACATTTTCAGTCTTCGAAC	1282
Db	362	CTTTCATCCCATTTGAAGAGTACATGTTGCATGTTGAAGAACATTTTCAGTCTTCGCAC	421
Qy	1283	G 1283	
Db	422	G 422	

```

RESULT 9
US-09-971-773-3
; Sequence 3, Application US/09971773
; Publication No. US20030115614A1
; GENERAL INFORMATION:
; APPLICANT: Yutaka KANDA
; APPLICANT: Mitsuo SATOH
; APPLICANT: Kazuyasu NAKAMURA
; APPLICANT: Kazuhisa UCHIDA
; APPLICANT: Toyohide SHINKAWA
; APPLICANT: Naoko YAMANE
; APPLICANT: Motoo YAMASAKI
; APPLICANT: NO. US20030115614a1u HANAI
; TITLE OF INVENTION: ANTIBODY COMPOSITION-PRODUCING CELL
; FILE REFERENCE: 249-202
; CURRENT APPLICATION NUMBER: US/09/971,773
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: JP 2000-308526
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: US 60/268,926
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 9196
; TYPE: DNA
; ORGANISM: Cricetus griseus
US-09-971-773-3

```

Query Match	15.0%;	Score 302;	DB 11;	Length 9196;
Best Local Similarity	100.0%;	Pred. No. 3.3e-76;		
Matches 302;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy 1	AACAGAAACCTATTCTCTGTGCGCTAACTAGAACCCAGAGTACAAATGTTTCCAAATTCCT	60		
Db 2450	AACAGAAACCTATTCTCTGTGCGCTAACTAGAACCCAGAGTACAAATGTTTCCAAATTCCT	2509		
Qy 61	TGACCTCCGAGAACAGACAGAGGGAGTTGAAACTCTGAAAAATGCGGGGCATGGACTGGTTC	120		
Db 2510	TGACCTCCGAGAACAGACAGAGGGAGTTGAAACTCTGAAAAATGCGGGGCATGGACTGGTTC	2569		
Qy 121	TGGCGTTGGATATAGCTCATCTCTTTTTCCTGGGGGACCTTATGTGTTTATATAGGTGGT	180		
Db 2570	TGGCGTTGGATATAGCTCATCTCTTTTTCCTGGGGGACCTTATGTGTTTATATAGGTGGT	2629		
Qy 181	CATTGTGTTCGAGATAATGACCCCTGACCATTTCTAGCAGAGAACTCTCCAGATTCCT	240		
Db 2630	CATTGTGTTCGAGATAATGACCCCTGACCATTTCTAGCAGAGAACTCTCCAGATTCCT	2689		
Qy 241	GCAAGCTGGAGCGCCTTAAAAACAAACAAATGAAGACTTGGAGGAAATGGGTGAGTCTCTC	300		
Db 2690	GCAAGCTGGAGCGCCTTAAAAACAAACAAATGAAGACTTGGAGGAAATGGGTGAGTCTCTC	2749		
Qy 301	CG 302			
Db 2750	CG 2751			

RESULT 10  
US-09-864-761-13359  
; Sequence 13359, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn. Sharron G.

```

1  ; APPLICANT: Rank, David R.
2  ; APPLICANT: Hanzel, David K.
3  ; APPLICANT: Chen, Wensheng
4  ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
5  ; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
6  ; FILE REFERENCE: Aeomica-X-1
7  ; CURRENT APPLICATION NUMBER: US/09/864,761
8  ; CURRENT FILING DATE: 2001-05-23
9  ; PRIOR APPLICATION NUMBER: US 60/180,312
10 ; PRIOR FILING DATE: 2000-02-04
11 ; PRIOR APPLICATION NUMBER: US 60/207,456
12 ; PRIOR FILING DATE: 2000-05-26
13 ; PRIOR APPLICATION NUMBER: US 09/632,366
14 ; PRIOR FILING DATE: 2000-08-03
15 ; PRIOR APPLICATION NUMBER: GB 24263.6
16 ; PRIOR FILING DATE: 2000-10-04
17 ; PRIOR APPLICATION NUMBER: US 60/236,359
18 ; PRIOR FILING DATE: 2000-09-27
19 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
20 ; PRIOR FILING DATE: 2001-01-30
21 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
22 ; PRIOR FILING DATE: 2001-01-30
23 ; PRIOR APPLICATION NUMBER: PCT/US01/00664
24 ; PRIOR FILING DATE: 2001-01-30
25 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
26 ; PRIOR FILING DATE: 2001-01-30
27 ; PRIOR APPLICATION NUMBER: PCT/US01/00665
28 ; PRIOR FILING DATE: 2001-01-30
29 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
30 ; PRIOR FILING DATE: 2001-01-30
31 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
32 ; PRIOR FILING DATE: 2001-01-30
33 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
34 ; PRIOR FILING DATE: 2001-01-30
35 ; PRIOR APPLICATION NUMBER: PCT/US01/00661
36 ; PRIOR FILING DATE: 2001-01-30
37 ; PRIOR APPLICATION NUMBER: PCT/US01/00670
38 ; PRIOR FILING DATE: 2001-01-30
39 ; PRIOR APPLICATION NUMBER: US 60/234,687
40 ; PRIOR FILING DATE: 2000-09-21
41 ; PRIOR APPLICATION NUMBER: US 09/608,408
42 ; PRIOR FILING DATE: 2000-06-30
43 ; PRIOR APPLICATION NUMBER: US 09/774,203
44 ; PRIOR FILING DATE: 2001-01-29
45 ; NUMBER OF SEQ ID NOS: 49117
46 ; SOFTWARE: Annotmax Sequence Listing Engine vers. 1.1
47 ; SEQ ID NO 13359
48 ; LENGTH: 503
49 ; TYPE: DNA
50 ; ORGANISM: Homo sapiens
51 ; FEATURE:
52 ; OTHER INFORMATION: MAP TO AL109847.3
53 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
54 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
55 ; US-09-864-761-13359

```



Db 186 TTGGTGGCTGGAATCATCTGGATGGCTATTCTAAAGGTGTCAACAGGAATTTGGGA 245  
Qy 1745 AAACAGGCTGTACCTTCTCAAGTCCGAGAGAGATAGAAACAGTCAAAATACCCTA 1804  
Db 246 GGACGGGCTATATCTCTCTCAAGTTCGAGAGAGATAGAAACGGTCAAGTACCCCA 305  
Qy 1805 CATATCTCTGAAGCTGAAATATAGATGGAGTGTAAAGAGATTAAACAACAGAAATTTAGTTC 1864  
Db 306 CATATCTCTGAGCTGAGAAATTAAGCTCAGATGGAGAGATAAACGACCAAACTCAGTTC 365  
Qy 1865 AGACCATCTCAG 1876  
Db 366 GACCAAACTCAG 377

RESULT 11  
US-09-864-761-13292  
; Sequence 13292, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aesomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 13292  
; LENGTH: 551  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:

; OTHER INFORMATION: MAP TO AL109847.3  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3  
US-09-864-761-13292  
Query Match 10.9%; Score 218.8; DB 9; Length 551;  
Best Local Similarity 81.6%; Pred. No. 9.5e-53;  
Matches 253; Conservative 0; Mismatches 57; Indels 0; Gaps 0;  
Qy 872 TGTTTAGACCTGTAAAGTGAGACATGCACAGACAGGTCTGGCCTCTCCACTGGACACTGGT 931  
Db 116 TGTGAATGGTGATGTATAGGAATACCATGTGTGTAATATATATATTTCAATATGT 175  
Qy 932 CAGGTGAAGTGAAGGACAAAATGTTCAAGTGTGAGCTCCCATTTGTAGACAGCTCC 991  
Db 176 CAGGTGAAGTGAAGGACAAAATGTTCAAGTGTGAGCTCCCATTTGTAGACAGCTCC 235  
Qy 992 ATCTCTGCTCTCTTACTTACCCTTGGCTGTACAGAGACCTTGCAGATCGACTCTCGA 1051  
Db 236 ATCCCGCTCTCCATATTTACCTTGGCTGTACAGAGACCTTGCAGATCGACTTGTAC 295  
Qy 1052 GAGTCCATGGTGATCCTGCACTGTGTGGGTATCCCAAGTTGTCAAATACTTTGATCCGTC 1111  
Db 296 GAGTGCATGGTGACCTTGCAGTGTGGGTGCTCTCAGTTTGTCAAATACTTTGATCCGCC 355  
Qy 1112 CACAACCTTGCTGGAAGGAAATAGAGAAACCAAGAAAGCTTGGCTTCAAAATC 1171  
Db 356 CACAGCCTTGCTAGAAAAGAAATAGAGAAAGCCACCAAGAGCTTGGCTTCAAAATC 415  
Qy 1172 CAGTTATTGG 1181  
Db 416 CAGTTATTGG 425  
RESULT 12  
US-09-864-761-18181  
; Sequence 18181, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aesomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 13292  
; LENGTH: 551  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:

```
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 18181
/ LENGTH: 248
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AL109847.1
/ OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.2
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5
/ OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6
/ OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.6
/ OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5
/ OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.1
/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7
/ OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.9
/ OTHER INFORMATION: SWISSPROT HIT: Q9VLZ7, EVALUE 6.50e+00
/ OTHER INFORMATION: NT HIT: AF038280.1, EVALUE 0.00e+00
/ OTHER INFORMATION: EST_HUMAN HIT: AW387766.1, EVALUE 0.00e+00
US-09-864-761-18181

Query Match 10.6%; Score 213.4; DB 9; Length 248;
Best Local Similarity 91.5%; Pred. No. 2.1e-51;
Matches 226; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 935 GTGAAGTGAAGACAAAATGTTCAAGTGTGTCAGAGCTCCCATTTGTACAGAGCTTCCATC 994
DB 1 GTGAAGTGAAGACAAAATGTTCAAGTGTGTCAGAGCTTCCCATTTGTACAGAGCTTCCATC 60

QY 995 CTCGTCTCTTACTTACCTTGGCTGTACCAAGAGACTTGCAGATCGACTCTCTGAGAG 1054
DB 61 CCCGTCTCCATATTACCTTGGCTGTACCAAGAGACTTGCAGATCGACTTGTACGAG 120

QY 1055 TCCATGGTGATCTCGCATGTGGTGGGTATCCAGTTTGTCAAATACTTGTATCGGTCCAC 1114
DB 121 TGCATGGTGATCCCTGCAGTGTGGTGGGTGTCTCAGTTTGTCAAATACTTGTATCGGTCCAC 180

QY 1115 AACCTTGGCTGAAGGGAATAGAGAAACCAAGAGCTTGGCTTCAACATCCAG 1174
DB 181 AGCCTTGGCTAGAAAAGAAATAGAGAAGCCACCAAGAGCTTGGCTTCAACATCCAG 240

QY 1175 TTATTGG 1181
DB 241 TTATTGG 247

RESULT 13
US-09-864-761-30284
/ Sequence 30284, Application US/09864761
/ Patent No. US20020048763A1
/ GENERAL INFORMATION:
/ APPLICANT: Penn, Sharon G.
/ APPLICANT: Rank, David R.
/ APPLICANT: Hanzel, David K.
/ APPLICANT: Chen, Wensheng
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
/ FILE REFERENCE: Aomic-X-1
/ CURRENT APPLICATION NUMBER: US/09/864,761
/ CURRENT FILING DATE: 2001-05-23
```

```
/ PRIOR APPLICATION NUMBER: US 60/180,312
/ PRIOR FILING DATE: 2000-02-04
/ PRIOR APPLICATION NUMBER: US 60/207,456
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/632,366
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: GB 24263.6
/ PRIOR FILING DATE: 2000-10-04
/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/00666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00664
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00669
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00665
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 30284
/ LENGTH: 248
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AL109847.3
/ OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
/ OTHER INFORMATION: SWISSPROT HIT: Q9VLZ7, EVALUE 6.50e+00
/ OTHER INFORMATION: EST_HUMAN HIT: AW387766.1, EVALUE 0.00e+00
/ OTHER INFORMATION: NT HIT: AF038280.1, EVALUE 0.00e+00
US-09-864-761-30284

Query Match 10.6%; Score 213.4; DB 9; Length 248;
Best Local Similarity 91.5%; Pred. No. 2.1e-51;
Matches 226; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 935 GTGAAGTGAAGACAAAATGTTCAAGTGTGTCAGAGCTCCCATTTGTACAGAGCTTCCATC 994
DB 1 GTGAAGTGAAGACAAAATGTTCAAGTGTGTCAGAGCTTCCCATTTGTACAGAGCTTCCATC 60

QY 995 CTCGTCTCTTACTTACCTTGGCTGTACCAAGAGACTTGTGCAGATCGACTCTCTGAGAG 1054
DB 61 CCCGTCTCCATATTACCTTGGCTGTACCAAGAGACTTGCAGATCGACTTGTACGAG 120

QY 1055 TCCATGGTGATCTCGCATGTGGTGGGTATCCAGTTTGTCAAATACTTGTATCGGTCCAC 1114
DB 121 TGCATGGTGATCCCTGCAGTGTGGTGGGTGTCTCAGTTTGTCAAATACTTGTATCGGTCCAC 180

QY 1115 AACCTTGGCTGAAGGGAATAGAGAAACCAAGAGCTTGGCTTCAACATCCAG 1174
DB 181 AGCCTTGGCTAGAAAAGAAATAGAGAAGCCACCAAGAGCTTGGCTTCAACATCCAG 240

QY 1175 TTATTGG 1181
DB 241 TTATTGG 247
```

```
Db      241 TTATTGG 247

RESULT 14
US-10-125-968-274
; Sequence 274, Application US/10125968
; Publication NO. US20030215805A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Palermo, Adam
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; APPLICANT: Elias, Josh
; APPLICANT: Mertens, Maureen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-032
; CURRENT APPLICATION NUMBER: US/10/125,968
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,163
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 1417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 274
; LENGTH: 481
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 204, 266, 320, 328, 402, 424, 462, 470
; OTHER INFORMATION: n = A,T,C or G
US-10-125-968-274

Query Match      7.4%; Score 149; DB 13; Length 481;
Best Local Similarity 84.9%; Pred. No. 1.8e-32;
Matches 213; Conservative 0; Mismatches 32; Indels 6; Gaps 4;

Qy      338 CAGGAAGAGTCGCTGTTTAGAAGAACACAGCTGTTTAAAGCCAAAGACAG-ATTGAAAAAT 396
Db      19  CGGCGAGGTACGCGTTTACAAGAGAGCAGCTGTTTAAAGCCAAAGACAGATTGAAAAAT 78

Qy      397 TACAAGAAACAGCTAGG--AATGATCTGGGAAGAGCATGAATCTTAAGGAGGAGGA 454
Db      79  TACAAGAAACAGACAGCAGTAATGCTCTGGGAAGGATCATGAATCTGAGGAGGAGGA 138

Qy      455 TTGAATAATGGAGCTAAAGAGCTCTGGTTTCTTCTACAAAGTCAATTGAAGAAATTAAGA 514
Db      139 TTGAATAATGGAGCTAAAGAGCTCTGGTTTCTTCTACAGAGTGAATTGAAGAAATTAAGA 198

Qy      515 AATTAGAAGGAACGAACCTCCAAAGACATGC-AGATGAATCTTTTGG--ATTTAGGAC 571
Db      199 ACTTANAGGAATGAATCTCCAAAGACATGCAAGATGAATTTCTTTGGATTATTAGGAC 258

Qy      572 ATCATCAAGG 582
Db      259 ATCATGANAAG 269

RESULT 15
US-09-864-761-1421
; Sequence 1421, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
```

```
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 1421
; LENGTH: 384
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL109847.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.9
US-09-864-761-1421

Query Match      6.8%; Score 136.4; DB 9; Length 384;
Best Local Similarity 76.6%; Pred. No. 7.4e-29;
Matches 167; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy      872 TGTTTAGACCTGTAAGTGAGACATGCACAGACAGCTGTGGCTCTCCACTGGACACTGGT 931
Db      167 TGTTCGAATGGTGGATGTATAGGAATACCATGTGTGTAATGATTATATGTTTCAATATTGT 226

Qy      932 CAGGTGAAGTGAAGCAACAAAATGTTCAAGTGTGTCAGCTCCCATTTGTAGACAGCTCC 991
Db      227 CAGGTGAAGTGAAGCAACAAAATGTTCAAGTGTGTCAGCTTCCTCCATTTGTAGACAGTCTTC 286

Qy      992 ATCTCGTCTCTCTTACTTACTCCCTTGGCTGTACAGAGACCTTGCAGATCGACTCTGTA 1051
Db      287 ATCCCGCTCTCCATATTTACCTTGGCTGTACAGAGACCTTCGAGATCGACTTGTATC 346

Qy      1052 GAGTCCATGGTGTATCTTCGAGTGTGGTGGGTATCCCGAG 1089
```

Db 347 GAGTGCATGGTGACCCCTGCAGTGGTGGTGCTCTCAG 384

Search completed: February 2, 2004, 19:02:53  
Job time : 717.477 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM\_nucleic - nucleic search, using sw model

Run on: February 2, 2004, 11:54:20 ; Search time 4312.69 Seconds  
(without alignments)  
11316.242 Million cell updates/sec

Title: US-09-971-773-1

Perfect score: 2008

Sequence: 1 aacagaactatttctctg.....gctgtgcctcaagcccatg 2008

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba.\*

2: em\_estbm.\*

3: em\_estcin.\*

4: em\_estmu.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_htc.\*

9: gb\_estcl.\*

10: gb\_est2.\*

11: gb\_htc.\*

12: gb\_est3.\*

13: gb\_est4.\*

14: gb\_est5.\*

15: em\_estfun.\*

16: em\_estom.\*

17: em\_gss\_hum.\*

18: em\_gss\_inv.\*

19: em\_gss\_pln.\*

20: em\_gss\_vrt.\*

21: em\_gss\_fun.\*

22: em\_gss\_mam.\*

23: em\_gss\_mus.\*

24: em\_gss\_pro.\*

25: em\_gss\_rtd.\*

26: em\_gss\_phg.\*

27: em\_gss\_vrl.\*

28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1809.4	90.1	2766	11 AK051811	AK051811 Mus muscu
2	1809.4	90.1	3052	11 AK048520	AK048520 Mus muscu
3	831.6	41.4	937	13 B0892101	B0892101 AGENCOURT
4	810	40.3	1066	12 BM928521	BM928521 AGENCOURT

5	768	38.2	940	13	BQ653329	BQ653329 AGENCOURT
6	757.6	37.7	900	13	BX452321	BX452321 AGENCOURT
7	751.2	37.4	886	13	BQ423860	BQ423860 AGENCOURT
8	741	36.9	923	13	BX429921	BX429921 AGENCOURT
9	741	36.9	1027	13	BX450010	BX450010 AGENCOURT
10	738	36.8	855	13	BUI84296	BUI84296 AGENCOURT
11	730.2	36.4	898	14	CB194828	CB194828 AGENCOURT
12	699.2	34.3	834	9	AUI124128	AUI124128 AGENCOURT
13	641.2	31.9	937	13	BUI51362	BUI51362 AGENCOURT
14	634.6	31.6	957	13	BX414472	BX414472 AGENCOURT
15	634.6	31.6	790	14	CB989769	CB989769 AGENCOURT
16	632	31.5	752	13	BQ604588	BQ604588 MI-P-CP1-
17	625.4	31.1	965	12	B1546364	B1546364 60318856
18	623.2	31.0	692	14	CB105611	CB105611 K-EST0130
19	623.2	31.0	692	14	CB105621	CB105621 K-EST0130
20	612	30.5	717	14	CB322023	CB322023 UI-CF-PN0
21	604.6	30.1	970	13	BQ926824	BQ926824 AGENCOURT
22	601.6	30.0	656	13	BQ552954	BQ552954 H4019310-
23	599.4	29.9	702	12	BG973221	BG973221 602842563
24	593.6	29.6	805	14	CA489662	CA489662 AGENCOURT
25	582.2	29.0	936	10	BE893107	BE893107 601437002
26	572.2	28.7	646	14	CB105920	CB105920 K-EST0135
27	573.6	28.6	769	10	BG501191	BG501191 602546241
28	567.8	28.3	731	13	BX456265	BX456265 BX456265
29	566.6	28.2	637	14	CB105617	CB105617 K-EST0130
30	563.4	28.1	613	14	CA539475	CA539475 C0279008-
31	559.6	27.9	622	14	CB105602	CB105602 K-EST0130
32	559	27.8	715	12	BG771085	BG771085 602719363
33	558.6	27.8	621	14	CB105873	CB105873 K-EST0135
34	555.6	27.7	618	14	CB105618	CB105618 K-EST0130
35	551	27.4	938	13	BX433236	BX433236 BX433236
36	548.8	27.3	608	14	CB105918	CB105918 K-EST0135
37	547.8	27.3	607	14	CB105608	CB105608 K-EST0130
38	547.8	27.3	607	14	CB105882	CB105882 K-EST0135
39	547.6	27.3	900	13	BQ731540	BQ731540 AGENCOURT
40	546.2	27.2	607	14	CB105899	CB105899 K-EST0135
41	544.8	27.1	653	10	BG619839	BG619839 602618014
42	540	26.9	610	14	CB105867	CB105867 K-EST0135
43	528.4	26.3	586	14	CB105886	CB105886 K-EST0135
44	526	26.2	585	14	CB105871	CB105871 K-EST0135
45	524	26.1	573	9	AI122135	AI122135 udl5d10.r

## ALIGNMENTS

RESULT 1

AK051811

LOCUS

DEFINITION

AK051811

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AK051811 2766 bp mRNA linear HTC 05-DEC-2002  
Mus musculus 12 days embryo eyeball cDNA, RIKEN full-length  
enriched library, clone:D330003K17 product:fucoyltransferase 8,  
full insert sequence.

AK051811 GI:26342229  
HTC; CAP trapper.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
99279253  
10349636  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20499374  
11042159

REFERENCE  
AUTHORS

3 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,  
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,  
Sumi,N., Iehi,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,  
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,  
Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,S., Watahiki,M.,  
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,  
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

4 Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,  
Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,  
Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,  
Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,K.,  
Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,  
Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,  
Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,  
Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,  
Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,  
Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,  
Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Buit,C.,  
Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,  
Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,  
Marchionni,L., Mashima,J., Mazzarelli,J., Mombassi,P., Nordone,P.,  
Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,  
Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,  
Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,D.,  
Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.  
and Hayashizaki,Y.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)  
21085660  
11217851

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

5 The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 2766)

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,  
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,  
Hayashida,K., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T.,  
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,  
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,  
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,  
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,  
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,  
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,  
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,P., Takaku-Akahira,S.,  
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,  
Muramatsu,M. and Hayashizaki,Y.  
Direct Submision

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216)

## COMMENT

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL:http://genome.gsc.riken.go.jp/  
URL:http://fantom.gsc.riken.go.jp/  
Location/Qualifiers

## FEATURES

## source

1. .2766  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM,DB:D230003K17"  
/db\_xref="taxon:10090"  
/clone="D230003K17"  
/tissue\_type="eyeball"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="12 days embryo"  
473 - 2200  
/note="unnamed protein product; fucosyltransferase 8  
(MGD|MG1:1858901, GB|NM\_016893, evidence: BLASTN, 99%,  
match=2175)  
putative"  
/codon\_start=1  
/protein\_id="BAC34777.1"  
/db\_xref="GI:26342230"  
/translation="MRAWTGSRWIMILFAMGTLTYIGHLVRDNDHPDHSRELS  
KILAKLERKQNEEDLRMAESLRIPGPIDQGTGRVRLVEQLVKAKEIQENYKK  
QARNLGDHEILRRITENGAKELWFFLQSELKKLEGLNEQLQHADELILLDGHHE  
RSIMTLYLSQTDGAGDREKAKOLTELVRRTIYLPNPKDCSKARKLVNKNKGC  
GYCOLHVVYCFMIAVGTORTLILESONWRYATGWTFRPVSSETCDRSGLSTGH  
WSBENDKNQVVELPIVDSLHPRPPYLPVLAPEDLADRLLRVHGDPAVWVSQFVKY  
LIRPQWLEKEIEETKKLGGKPVIGVHVRRTDKVGEARAFHFIETWVHEHFQL  
LARMQVKRVYLATDDPTLLKEAKTYSNYSFISNSISWSAGLHNRITENSLRGV  
ILDTHISQADFLVCTPSSQVCRVAYEIMQTLHPDASANFHSLLDIYFGQNAHQI  
AVYHPKRTETEEIPMEPCDIIGVAGNHWGYSKGNRKLKGTGLYPSYKVKREKLETVK  
YPTYPEAEK"  
2750 .2755  
/note="putative"  
2766

## CDS

BASE COUNT 838 a 591 c 619 g 718 t  
ORIGIN  
Query Match 90.1%; Score 1809.4; DB 11; Length 2766;  
Best Local Similarity 94.1%; Pred. No. 0;  
Matches 1891; Conservative 0; Mismatches 116; Indels 2; Gaps 1;  
Qy 1 AACAGAACTATTTTCTGTGTGCTAACTAGAACACAGAGTACAATGTTTCCAAATCTT 60  
Db 372 AAAAGAACCTATTTTCTGTGTGCTAACTAGAACACAGAGTACAATGTTTCCAGTCTT 431  
Qy 61 TGAGCTCCGAGAACACAGA--AGGAGTGTGAACTCTGAAATCGGGGCATGGACTGTT 118  
Db 432 TGAGCTCCGAGAACAGACAGAGTTGAACTCTGAAATCGGGGCATGGACTGTT 491  
Qy 119 CCTGGCGTTGGATTATGCTCAATCTTTTCTGGGGGACCTTATGTTTTATATAGTG 178  
Db 492 CCTGGCGTTGGATTATGCTCAATCTTTTCTGGGGGACCTTATGTTTTATATAGTG 551  
Qy 179 GTCAATTTGTTTCAGATAATGACACACCTGACCATTTCTAGCAGAGAACTCTCCAAGATTC 238  
Db 552 GTCAATTTGTTTCAGATAATGACACACCTGACCATTTCTAGCAGAGAACTCTCCAAGATTC 611  
Qy 239 TTGCAAGCTGGAGCGCTTAAACCAACAAATGAAGACTTTGAGGAGAAATGGCTGAGTCTC 298  
Db 612 TTGCAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTTGAGGCGAATGGCTGAGTCTC 671  
Qy 299 TCCGAATACCAAGAGGCCCTTATTGATCAGGGGACAGCTACAGGAAGAGTCCGCTGTTTAG 358  
Db 672 TCCGAATACCAAGAGGCCCTTATTGATCAGGGGACAGCTACAGGAAGAGTCCGCTGTTTAG 731  
Qy 359 AAGAACAGCTGTTTAAAGCCCAAGAACAGATTGAAATTTCAAGAAACAAGCTAGGAATG 418  
Db 732 AAGAACAGCTGTTTAAAGCCCAAGAACAGATTGAAATTTCAAGAAACAAGCTAGGAATG 791  
Qy 419 ATCTGGAAAGGAGTATCAATAATCTTAAGGAGGAGGATTGAAATGGAGCTCT 478  
Db 792 GTCTGGGGAAGGATCATGAATCTTTAAGAGGAGGATTGAAATGGAGCTCT 851  
Qy 479 GGTTTTTCTACAAAGTGAATTTGAAGAAATTAAGAAATTTAGAAAGAAACGAACCTCAA 538

## polyA\_signal

2750 .2755  
/note="putative"  
2766

## polyA\_site

2750 .2755  
/note="putative"  
2766

Db 852 GGTGTTTCTACAAAGCGAACTGAAGAAATTAAGCAATTAAGAGAAATGAATCTCCAAA 911  
Qy 539 GACATGACAGATGAATCTTTTGGATTTAGACATCATGAAGTCTATCATGACAGATC 598  
Db 912 GACATGACAGATGAATCTTTTGGATTTAGACATCATGAAGTCTATCATGACAGATC 971  
Qy 599 TATATCTAGTCTCAGTCAAAACAGATGAGCAGAGTGTAGTGGCGGGAAGAAAGCCAAAGATC 658  
Db 972 TATATCTAGTCTCAGTCAAAACAGATGAGCAGAGTGTAGTGGCGGGAAGAAAGCCAAAGATC 1031  
Qy 659 TGACAGAGTGTCTCAGCGGGAATTAATATCTGACAGAAATCTCCAAAGACCTGACGAAAG 718  
Db 1032 TGACAGAGTGTCTCAGCGGGAATTAATATCTGACAGAAATCTCCAAAGACCTGACGAAAG 1091  
Qy 719 CCAGAAAGTGGTATGTAATCAACAAGGCTGTGGCTATGGATGTCATCTCCATCATG 778  
Db 1092 CCAGAAAGTGGTATGTAATCAACAAGGCTGTGGCTATGGATGTCATCTCCATCATG 1151  
Qy 779 TGGTTTACTGCTTCATGATGCTTATGGCACCAGCGAACTCATCTTTGGAATCTCAGA 838  
Db 1152 TGGTCTACTGTTTCAATGCTTATGGCACCAGCGAACTCATCTTTGGAATCTCAGA 1211  
Qy 839 ATTGGCGTCTGCTACTGAGAGATGGGAGACTGTGTTTGTAGACCTGTAAGTGAACATGCA 898  
Db 1212 ATTGGCGTCTGCTACTGAGAGATGGGAGACTGTGTTTGTAGACCTGTAAGTGAACATGTA 1271  
Qy 899 CAGACAGTCTGGGCTCTCCACTGACACTGGTCAAGTGAAGTGAAGGACAAATGTTTC 958  
Db 1272 CAGACAGTCTGGGCTCTCCACTGACACTGGTCAAGTGAAGTGAAGGACAAATGTTTC 1331  
Qy 959 AAGTGGTCCGAGTCTCCCTCTGAGACAGCTCCATCTCGTCTCTTACTTACCTCTGG 1018  
Db 1332 AAGTGGTCCGAGTCTCCCTCTGAGACAGCTCCATCTCGGCTCTTACTTACCAGTGG 1391  
Qy 1019 CTGTACCAAGAGACCTTCAGATCGACTCTCTGAGAGTCCATGGTGAATCCTCGAGTGGT 1078  
Db 1392 CTGTTCCAGAGACCTTCAGAGACCTCTTAAGAGTCCATGGTGAACCTCGAGTGGT 1451  
Qy 1079 GGGTATCCAGTGTGCAAACTGATCGCTGCAACACTTGGTGGTGAAGGGAATAG 1138  
Db 1452 GGGTATCCAGTGTGCAAACTGATCGCTGCAACACTTGGTGGTGAAGGGAATAG 1511  
Qy 1139 AAGAAACCAAGAGCTGGCTTCAAACTGAGTTCATGAGTCCATGTCAGACGCA 1198  
Db 1512 AAGAAACCAAGAGCTGGCTTCAAACTGAGTTCATGAGTCCATGTCAGACGCA 1571  
Qy 1199 CTGACAAAGTGGGAACAGAGCAGCTTCCATCCCATTTGAGGAATPACATGTCAGCGTTG 1258  
Db 1572 CAGACAAAGTGGGAACAGAGCAGCTTCCATCCCATTTGAGGAATPACATGTCAGCGTTG 1631  
Qy 1259 AAGAACTTTTTCAGCTTCGAAACAGAGTGAATGAATGAATGAATGAATGAATGAATGAAT 1318  
Db 1632 AAGAACTTTTTCAGCTTCGAAACAGAGTGAATGAATGAATGAATGAATGAATGAATGAAT 1691  
Qy 1319 CCATGATGACCTCTTTTGAAGGAGGCAAGACAAAGTACTCCAAATATGAATTTA 1378  
Db 1692 CTACTGATGATCTCTTTTGAAGGAGGCAAGACAAAGTACTCCAAATATGAATTTA 1751  
Qy 1379 TTAGTGAATCTCTTTTGGTCTGAGTGGACTACAAACCGATACACAGAAATTTAC 1438  
Db 1752 TTAGTGAATCTCTTTTGGTCTGAGTGGACTACAAACCGATACACAGAAATTTAC 1811  
Qy 1439 TTGGGGGCTGATCTCGATATACACTTTCTCTCCAGGCTGACCTTCTGTGTGTAAT 1498  
Db 1812 TTGGGGGCTGATCTCGATATACACTTTCTCTCAGAGCTGACCTTTCTAGTGTGTAAT 1871  
Qy 1499 TTTGATCCAGCTGTGAGGTTGCTTATGAATCATCAACACTGATCTCTGATGCT 1558  
Db 1872 TTTGATCCAGCTGTGAGGTTGCTTATGAATCATCAACACTGATCTCTGATGCT 1931  
Qy 1559 CTGCAAACTCTTCTTTAGATGACATCTACTATTTTGGAGGCGCAAAATGCCCAACACC 1618

Db 1932 CTGCGAACTTTCATTTCTTTGGATGACATCTACTATTTTGGAGGCCAAATCCCAACAATC 1991  
Qy 1619 AGATTGAGTATTATCTCTCACCAACTGAACTAAAGAGGAAATCCCAATGGAACCTGGAG 1678  
Db 1992 AGATTGCTTTTATCTCTCAACAACCTGAACTGAAGAGGAAATTTCCATGGAACCTGGAG 2051  
Qy 1679 ATATCATTTGGTGTGGCTGGAAACCATTTGGAATGTTTACTCTTAAAGGTCTCAACAGAAAC 1738  
Db 2052 ATATCATTTGGTGTGGCTGGAAACCATTTGGATGTTTATTTCTTAAAGGTATCAACAGAAAC 2111  
Qy 1739 TAGGAAAACAGGCTGTACCTCTTACAAAGTCCGAGAGAGATAGAAACAGTCAAT 1798  
Db 2112 TTGAAAACAGGCTGTATATCTCTTCAAAAGTCCGAGAGAGATAGAAACAGTCAAT 2171  
Qy 1799 ACCCTACATATCTTGAAGCTGAAATAGAGATGGAGTGAAGATTAACAAACAGAAAT 1858  
Db 2172 ATCCACATATCTTGAAGCTGAAATAGAGATGAAGTGAAGATTTAAACAGAACT 2231  
Qy 1859 TAGTTTCAGACCATCTTAGCCAAAGCAGAGAGCCAGACTTAAATATGTTTCTTTCAGACAG 1918  
Db 2232 CACTTCAGACCATCTTAGCCAAAGCAGAGAGCCAGACTTAAACAGTGGTTCATTTGATAGAC 2291  
Qy 1919 ATGCTCCGACCAAGAGCAAGTGGAAACCTCAGATGCTGCACTGGTGGAGCGCTCTTT 1978  
Db 2292 ACGTCCACCAAGAGCAAGCGGGAACCTCAGATGCTGCACTGGTGGAGCGCTCTTT 2351  
Qy 1979 GTGAAGGCTGCTGCTGCTCAAGCCCAT 2007  
Db 2352 ATGAAGGCTGTGTTACCTCAAGCCCAT 2380

RESULT 2  
AK048520  
LOCUS  
DEFINITION Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched library, clone:Cl:0068P22 product: fucosyltransferase 8, full insert sequence.  
ACCESSION AK048520  
VERSION AK048520.1 GI:26339371  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE 1  
AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
REFERENCE 2  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PUBMED 11042159  
REFERENCE 3  
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, K., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipillar sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE 20530913  
PUBMED 11076861  
REFERENCE 4  
AUTHORS Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,



Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehlin, P., Lewis, S., Mateu, Y., Nikaide, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Clercher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazza, R., Mombert, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wyshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S., and Hayashizaki, Y.

Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)  
21085660  
11217851

5  
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase 1 & II team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 3052)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.

Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/.

Location/Qualifiers  
1..3052  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CS7BL/6J"  
/db\_xref="FANTOM DB:C130068P22"  
/db\_xref="taxon:10090"  
/clone="C130068P22"  
/tissue\_type="head"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="16 days embryo"  
554..2381  
/note="unnamed protein product; fucosyltransferase 8 (MGD|MG1:1858901, GB|NM\_016893, evidence: BLASTN, 99%, match=2175)  
putative"

## CDS

/codon\_start=1  
/protein\_id="BAC33357.1"  
/db\_xref="GI:26339372"  
/translation="MRAWTGSRWIMILFAMWGTLLFYIGHLVLDNDHPHSHRELS  
KILAKRLKQONEDLRMAESLRIPGPIDQGTATGRVRVLEQLVKAKEQIENYKK  
QARNGLKDHILLRRRTENGAKLWFFLOSELKKLHLEGNLEORHAEIILLDGHHE  
RSIMTDLYLSQDTGAGDREKAKDLTELVRRI TYLONPKDCSKARKLVNLNKGK  
GYGCOLHHVYVCFMIAVGTORTLILLESQNRVATGGWETVPRPVSFETCDRSGLSTGH  
WSGEVDKNIQVVELPIVDSLHPRPYPPLAVPEDLADRLLRVHGDDPAWVWSQFVKY  
LIRPQWLKEIEEATKLGFGKPGVIGVHVRRTDKVGEAFHPIIEETMVHVEHFLQ  
LARMQVDKRVILATDPTLLKEAKTKYSNIEPISDNSISWSAGLHNRVYTESLGV  
ILDHFLSQADFLVCTFSVOCVAVYIMOTLHPDASANPHSLDDI VYFGQNAHNOI  
AVYHPKPEETEEIPMEPGDIIIGVAGNHWGVSKGINKLGTGLYPSYKVKREKJETVK  
IPTYPEAEK"  
polyA\_signal 3036..3041  
/note="putative"  
polyA\_site 3052  
/note="putative"  
BASE COUNT 895 a 685 c 684 g 788 t  
ORIGIN  
Query Match 90.1%; Score 1809.4; DB 11; Length 3052;  
Best Local Similarity 94.1%; Pred. No. 0;  
Matches 1891; Conservative 0; Mismatches 116; Indels 2; Gaps 1;  
QY 1 AAGAAACCTATTCTCTGTGTGCTAACTAGAACACAGAGTACAAATGTTTCCAAATCTT 60  
DB 553 AAAAGAAACCTATTCTCTGTGTGCTAACTAGAACACAGAGTACAAATGTTTCCAGTCTT 612  
QY 61 TGAGCTCCGAGAACACAGACA -- AGGAGTGTGAAACTCTGAAATCGGGCATCGACTGCTT 118  
DB 613 TGAGCTCCAGAGATAGAGACAGAGTTGAAATCTTGAAATCGGGCATCGACTGCTT 672  
QY 119 CCTGCGTGGATTATGCTCATCTCTTTTCTGCTGGGACCTTATTTGTTTATATAGGTG 178  
DB 673 CCTGCGTGGATTATGCTCATCTCTTTTCTGCTGGGACCTTGTATTTATATAGGTG 732  
QY 179 GTCAATTTGGTTCGAGATAATGACACACCTGACCAATCTTAGCAGAGAACTCTCCAAGATTC 238  
DB 733 GTCAATTTGGTTCGAGATAATGACACACCTGACCAATCTCTCCAGCAGAGAACTCTCCAAGATTC 792  
QY 239 TTGCAAGCTGGAGCGCTTAAACACAAATGAAGACTTCGAGGAGATGCTGAGTCTC 298  
DB 793 TTGCAAGCTTGAACGCTTAAACACAGCAAAATGAAGACTTCGAGGAGATGCTGAGTCTC 852  
QY 299 TCCGAATACCAAGAGGCGCTTATTTGATCAGGGGACAGCTTACAGGAGAGTCCGCTGTTTATG 358  
DB 853 TCCGAATACCAAGAGGCGCTTATTTGATCAGGGGACAGCTTACAGGAGAGTCCGCTGTTTATG 912  
QY 359 AAGAACAGCTGTTTAAAGCCAAAGACAGATTGAAATTTACAGAAACAACTAGGAATG 418  
DB 913 AAGAACAGCTGTTTAAAGCCAAAGACAGATTGAAATTTACAGAAACAACTAGGAATG 972  
QY 419 ATCTGGGAAAGGATCATGAATCTTTAAGGAGGAGTGTGAAATTTGAGACTTAAAGAGCTCT 478  
DB 973 GTCTGGGAAAGGATCATGAATCTTTAAGAGGAGATTTGAAATTTGAGACTTAAAGAGCTCT 1032  
QY 479 GGTTTTTTCTACAAAGTGAATTTGAAGAAATTTAAGAAATTTAAGAGGAAACGAACCTCCAAA 538  
DB 1033 GGTTTTTTCTACAAAGCGAACTGAAGAAATTTAAGAAATTTAAGAGGAAATGAACCTCCAAA 1092  
QY 539 GACATGAGATGAATTTCTTTTGGATTAGGACATCATGAAGGTCTATCATGACAGATC 598  
DB 1093 GACATGAGATGAATTTCTTTTGGATTAGGACATCATGAAGGTCTATCATGACAGATC 1152  
QY 599 TATCTACCTCAGTCAAAACAGATGAGCAGGTGAGTGGCGGAGAAAGAAAGCAAGATC 658  
DB 1153 TATCTACCTCAGTCAAAACAGATGAGCAGGTGAGTGGCGGAGAAAGAAAGCAAGATC 1212  
QY 659 TGACAGAGCTGGTCCAGCGGAGAAATAACATATCTGCAGAAATCCCAAGGATCGCAGCAAG 718  
DB 1213 TGACAGAGCTGGTCCAGCGGAGAAATAACATATCTCCAGAAATCCTTAAGGACTGCAGCAAG 1272

```
Qy 719 CCAGAAAGCTGTGTATGTAATATCAAAAGGCTGTGGCTATGGATGTCAACTCCATCATG 778
Db 1273 CCAGGAAGCTGGTGTGTAACATCAATAAGGCTGTGGCTATGGTGTGTAACCTCCATCAG 1332
Qy 779 TGGTTTACTGCTCTCATGATTTGTTATGGCAAGGAGGAGCACTCATCTTTGGAATCTCAGA 838
Db 1333 TGGTCTACTGTTTATGATTTGTTATGGCAAGGAGGAGCACTCATCTTTGGAATCTCAGA 1392
Qy 839 ATTGGCGCTATGCTACTCGAGAGTGGGAGACTGTGTTTATAGACCTGTAAAGTGAAGCATGCA 898
Db 1393 ATTGGCGCTATGCTACTCGAGAGTGGGAGACTGTGTTTATAGACCTGTAAAGTGAAGCATGTA 1452
Qy 899 CAGACAGGTCTGGCCCTCTCCACTGGACACTCGTTCAGAGTGAAGTGAAGGACAAAAATGTTTC 958
Db 1453 CAGACAGATCTGGCCCTCTCCACTGGACACTCGTTCAGAGTGAAGTGAAGGACAAAAATGTTTC 1512
Qy 959 AAGTGGTCGAGCTCCCAATTTGATAGACAGGCTCCATCTCTGCTCTCTTACTTACCTTGG 1018
Db 1513 AAGTGGTCGAGCTCCCAATTTGATAGACAGGCTCCATCTCTGCTCTCTTACTTACCTTGG 1572
Qy 1019 CTGTACCAAGAGACCTTCGACATCGACTCCCTGAGAGTCCATGGTGATCCTCGAGTGGT 1078
Db 1573 CTGTTTCCAGAGACCTTCGACAGGACTCCCTAAGAGTCCATGGTGACCTCGAGTGGT 1632
Qy 1079 GGGTATCCAGTTTGTCAAATCTTGTATCCGTCACAACTTTGGCTGGAAGGAAATAG 1138
Db 1633 GGGTCTCCAGTTTGTCAAATCTTGTATCCGTCACAACTTTGGCTGGAAGGAAATAG 1692
Qy 1139 AAGAAACCAACGAAGCTTGGTTCAAACATCCAGTTATTTGAGTCCATGTCAAGCA 1198
Db 1693 AAGAAAGCACCAGAAAGCTTGGTTCAAACATCCAGTTATTTGAGTCCATGTCAAGCA 1752
Qy 1199 CTGACAAAGTGGGACAGACAGCTTCCATCCATTTGAGATACATGTGACAGTTG 1258
Db 1753 CAGACAAAGTGGGACAGACAGCTTCCACCCATCGAGAGTACATGTGACAGTTG 1812
Qy 1259 AAGAAATTTTTCAGCTTCTCGAACCGAGATGAAGTGGATAAAAAAGAGTGTATCTGG 1318
Db 1813 AAGAAATTTTTCAGCTTCTCGACCGAGATGAAGTGGATAAAAAAGAGTGTATCTGG 1872
Qy 1319 CACTGTATGACCTTCTTTGTTTAAAGGAGGCAAGCAAGTACTTCCAAATATGAATTTA 1378
Db 1873 CTACTGTATGATCTCTATTTGTTTAAAGGAGGCAAGCAAGTACTTCCAAATATGAATTTA 1932
Qy 1379 TTAGTGTAACTCTATTTCTTGGTTCAGCTGAGCTACACAAACGATACACAGAAATTCAC 1438
Db 1933 TTAGTGTAACTCTATTTCTTGGTTCAGCTGAGCTACACAAACGATACACAGAAATTCAC 1992
Qy 1439 TTGGGGGGCTGATCCTGGATATACACTTTCTCTCCAGGCTGACTTCTCTGTGTGACTT 1498
Db 1993 TTGGGGGGTGTATCCTGGATATACACTTTCTCTACAGGCTGACTTCTGTGTGACTT 2052
Qy 1499 TTTTATCCAGGCTGTGATGGTGTGTTATGAATCATGCAACCACTGATCCTGATGCT 1558
Db 2053 TTTTATCCAGGCTGTGATGGTGTGTTATGAATCATGCAACCACTGATCCTGATGCT 2112
Qy 1559 CTGCAAACTTCATTTCTTATGATGATCATCTACTATTTTGGAGGCCAAATGCCACACAC 1618
Db 2113 CTGCAAACTTCATTTCTTATGATGATCATCTACTATTTTGGAGGCCAAATGCCACACATC 2172
Qy 1619 AGATTGCAAGTTTATCTCAACCACTCGAAGTAAAGAGGAAATCCCAATGAACTGGAG 1678
Db 2173 AGATTGCTGTTTATCTCAACCACTCGAAGTAAAGAGGAAATCCCAATGAACTGGAG 2232
Qy 1679 ATATCATTTGGTGTGGTGAACCAATTTGGAATGGTTACTCTAAAGGTTGTCAACAGAAAC 1738
Db 2233 ATATCATTTGGTGTGGTGAACCAATTTGGAATGGTTTATCTAAAGGTTGTCAACAGAAAC 2292
Qy 1739 TAGGAAAAACAGGCTGTATCTTCTACAAAGTCCGAGAGAGATGAAGACAGTCAAT 1798
Db 2293 TTGGAAGAAACAGGCTGTATCTTCTACAAAGTCCGAGAGAGATGAAGACAGTCAAGT 2352
Qy 1799 ACCCTACATATCTGAAAGCTGAAAGATAGAGATGGAGTGAAGATTTAAACACAGAAAT 1858
```

```
Db 2353 ATCCCATATCTCTGAAGCTCAAAATAGAGATGAAGTAGAAGATTAAACAGAACT 2412
Qy 1859 TAGTTTCAGACCATCTAGCCCAAGCAGAGACCCAGACTTAACATATGGTTTCATTCACAGAC 1918
Db 2413 CACTTCAGACCATCTAGCCCAAGCAGAGACCCAGACTTAACATATGGTTTCATTCATAGAC 2472
Qy 1919 ATGTCTCCGCCCAAGAGCAAGTGGGAAACCCCTCAAGTGTGACATGTGCTGGTGGAAACGCTCTTT 1978
Db 2473 ACCTTCACACACAGAGCAAGGGAACCCCTCAAGTGTGACATGTGCTGGTGGAAACGCTCTTT 2532
Qy 1979 GTGAAGGGCTGTCTGTGCTCCCTCAAGCCCAT 2007
Db 2533 ATGAAGGGCTGTGTGGTCCCTCAAGCCCAT 2561

RESULT 3
BQ892101
LOCUS
DEFINITION BQ892101 937 bp mRNA linear EST 16-AUG-2002
5', mRNA sequence.
ACCESSION BQ892101
VERSION BQ892101.1 GI:22284115
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 937)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Tissue Procurement: Susan L. Sullivan, Ph.D.
cDNA Library Preparation: ResGen, Invitrogen Corp
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13875 row: d column: 14
High quality sequence start: 130
High quality sequence stop: 780.
FEATURES
Location/Qualifiers
1..937
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6389725"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_129"
/note="Organ: olfactory epithelium; Vector:
pCW-SPOrt6.1.cdb; Site 1: EcoRV; Site 2: NotI; Cloned
unidirectionally. Primer: Oligo dt. Average insert size
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
is a NIH_MGC Library."
BASE COUNT 268 a 215 c 209 g 244 t 1 others
ORIGIN
Query Match 41.43; Score 831.6; DB 13; Length 937;
Best Local Similarity 94.03; Pred. No. 3.9e-187;
Matches 875; Conservative 0; Mismatches 55; Indels 1; Gaps 1;
Qy 835 CAGAATTTGGCGCTATGCTACTGGAGATGGGAGACTGTGTTTAGACCTGTAAAGTGAACA 894
Db 1 CAGAATTTGGCGCTATGCTACTGGAGATGGGAGACTGTGTTTAGACCTGTAAAGTGAACA 60
Qy 895 TGCAACAGACAGGCTGTGACCTCTCCACTGGACACTGGTCAAGTGAAGTGAAGCAAAAAT 954
Db 61 TGTAACAGACAGATCTGACCTCTCCACTGGACACTGGTCAAGTGAAGTGAAGCAAAAAC 120
Qy 955 GTTCAAGTGGTCGAGCTCCCATTTGTAGACAGCCTCCATCTCTCTTACTTACTCC 1014
```

|||||  
121 ATTCAAGTGGTGCAGCTCCCATTTAGACAGCTCCATCTCGGCCTCTTACTATCCA 180  
QY  
1015 TTGGCTGTACAGAAAGACCTTCAGATCGACTCCCTGAGAGTCCATGGTGTATCTTCAGTG 1074  
Db  
181 CTGGCTGTTCAGAAAGACCTTCAGACCGACTTCCTAAGAGTCCATGGTGTACCTGCACTG 240  
QY  
1075 TGGTGGGTATCCAGCTTTGTCAAAATATCTTGATTCGCTCCACACCTTGGCTGGAAAGGAA 1134  
Db  
241 TGGTGGGTGTCCAGTTTGTCAAAATATCTTGATTCGCTCCACACCTTGGCTGGAAAGGAA 300  
QY  
1135 ATAGAAGAAACACCAAGAGCTTGGCTTCAAAATATCTTGATTCGCTCCACACCTTGGCTGGAAAGGAA 1194  
Db  
301 ATAGAAGAAACACCAAGAGCTTGGCTTCAAAATATCTTGATTCGCTCCACACCTTGGCTGGAAAGGAA 360  
QY  
1195 GGCACCTGCAAAAGTGGGAACAGAGCGCTTCCATCCATTCAGAGTGAATAATCATGTTACAC 1254  
Db  
361 CCACAGACAAAGTGGGAACAGAGCGCTTCCATCCATTCAGAGTGAATAATCATGTTACAC 420  
QY  
1255 GTTGAAGAACATTTTCAGCTTCTCGAAGCGCAGAAATGAAAGTGGATGATAAAGAGTGTAT 1314  
Db  
421 GTTGAAGAACATTTTCAGCTTCTCGAAGCGCAGAAATGAAAGTGGATGATAAAGAGTGTAT 480  
QY  
1315 CTGGCCACTGATGACCTTCTTTGTTAAAGAGGAGGAAAGACAAAGTACTCCATTTATGAA 1374  
Db  
481 CTGGCTACTGATGATCTTCTTTGTTAAAGAGGAGGAAAGACAAAGTACTCCATTTATGAA 540  
QY  
1375 TTATTATAGTAACTCTATTTCTTTGGTTCAGCTGAGCTACACCAACCGATACACAGAAAT 1434  
Db  
541 TTATTATAGTAACTCTATTTCTTTGGTTCAGCTGAGCTACACCAACCGATACACAGAAAT 600  
QY  
1435 TCACCTTCGGGGCGGTGATCTCGATATACATTTTCTCTCCAGCTGAGCTTCTCTGTGTGT 1494  
Db  
601 TCACCTTCGGGGCGGTGATCTCGATATACATTTTCTCTCCAGCTGAGCTTCTCTGTGTGT 660  
QY  
1495 ACTTTTTCATCCAGCTGTGAGGTGTGTTAAGAAATCATGCAAACTGCAATCCTGAT 1554  
Db  
661 ACTTTTTCATCCAGCTGTGAGGTGTGTTAAGAAATCATGCAAACTGCAATCCTGAT 720  
QY  
1555 GCCTCTGCAAACTTCCATTTCTTTAGATGACATCTACTATTTTGGAGGCAAAATGCCAC 1614  
Db  
721 GCCTCTGCAAACTTCCATTTCTTTAGATGACATCTACTATTTTGGAGGCAAAATGCCAC 780  
QY  
1615 AACAGATTGCAATTTATCTCCACCAACCTCGAATTAAGAGGAAATCCCATGGAACCT 1674  
Db  
781 AATCAGATTGCTGTTTATCTCCACCAACCTCGAATTAAGAGGAAATCCCATGGAACCT 840  
QY  
1675 GGAGATATCATTTGCTGCTGGAACCT-ATTGGAATGTTTACTTAAAGGTGTCACAG 1733  
Db  
841 GGAGATATCATTTGCTGCTGGAACCTAATGGGATGTTTATTTCTAAAGGTATCAACAG 900  
QY  
1734 AAACTAGCAAAACAGGCTCTGACCTTCC 1764  
Db  
901 AAACTAGCAAAACAGGCTTATATCCCTTC 931

RESULT 4  
BM928521  
LOCUS  
DEFINITION BM928521 1066 bp mRNA linear EST 12-MAR-2002  
5', mRNA sequence.  
AGENCOURT\_6715254 NIH\_MGC\_100 Homo sapiens cdna clone IMAGE:5798262  
ACCESSION  
BM928521  
VERSION  
BM928521.1 GI:19378900  
KEYWORDS  
EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 1066)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.

Email: csapbe-re@mail.nih.gov  
Tissue Procurement: CGAP (Stanford)  
cdna Library Preparation: Rubin Laboratory  
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LiCM2025 row: p column: 07  
High quality sequence stop: 666.

FEATURES  
Location/Qualifiers  
1..1066  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clones="IMAGE:5798262"  
/tissue\_type="hepatocellular carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_100"  
/notes="Organ: liver; Vector: pOTB7; Site: 1: XhoI; Site 2:  
EcoRI; cdna made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCACGAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."  
BASE COUNT 315 a 235 c 248 g 264 t 4 others  
ORIGIN

Query Match 40.3%; Score 810; DB 12; Length 1066;  
Best Local Similarity 90.7%; Pred. No. 5.7e-182;  
Matches 894; Conservative 0; Mismatches 89; Indels 3; Gaps 3;  
QY 465 AGCTAAAGAGCTCTGGTTTCTTACAAAGTGAATTCGAAGAAATTAAGAATAATAGAGG 524  
Db 1 AGCTAAAGAGCTCTGGTTTCTTACAGAGTGAATTCGAAGAAATTAAGAATAATAGAGG 60  
QY 525 AAACGAACCTCAAGACATGCAGATGAATTTCTTTGGATTTAGGACATCATGAAGGTC 584  
Db 61 AATGAACTCAAGACATGCAGATGAATTTCTTTGGATTTAGGACATCATGAAGGTC 120  
QY 585 TATCATGACAGATCTATCTACCTCAGTCAAAAGATGGAGCAGGTGAGTGGCGGAAAA 644  
Db 121 TATAATGCGGATCTATCTACCTCAGTCAAGACATGGAGCAGGTGATTTGGCGGAAAA 180  
QY 645 AGAGCCAAAGATCTGACAGAGCTGGTCCAGCGGAGAAATACATATCTCGAGAATCCCAA 704  
Db 181 AGAGCCAAAGATCTGACAGAACTGGTTTCAGCGGAGAAATACATATCTTCAGAATCCCAA 240  
QY 705 GGACTGACGAAAGCCAGAGAGCTGGTATGTAATATCAACAAAGGCTGTGGCTATGGATG 764  
Db 241 GGACTGACGAAAGCCAAAGAGCTGGTATGTAATATCAACAAAGGCTGTGGCTATGGCTG 300  
QY 765 TCAACTCCATCATGTGGTTTACTCTCTTCATGATTTATGGACCCAGCGAACTCAT 824  
Db 301 TCAGCTCCATCATGTGGTCTACTCTCTTCATGATTTATGGACCCAGCGAACTCAT 360  
QY 825 CTTGGAATCTCAGAAATTTGGCGCTATGCTACCTGGAGGATGGAGACTGTGTTTAGACCTGT 884  
Db 361 CTTGGAATCTCAGAAATTTGGCGCTATGCTACCTGGGATGGAGACTGTGTTTAGACCTGT 420  
QY 885 AAGTGAGACATGCACAGACAGCTCTGGCTCTCCACTGGACACTGGTTCAGGTGAAGTGA 944  
Db 421 AAGTGAGACATGCACAGACAGATCTGGCATCTCCACTGGACACTGGTTCAGGTGAAGTGA 480  
QY 945 GGACAAATAATTTCAAGTGTGAGCTCCCATTTGTAGACAGCTCCATCTCTGCTCTCC 1004  
Db 481 GGACAAATAATTTCAAGTGTGAGCTCCCATTTGTAGACAGCTCTTCATCTCCCGCTCTCC 540  
QY 1005 TTACTTACCTTGGCTGTGTACCAGAGACCTTTCAGATGCTCCTCTGAGAGTCCATGGTGA 1064

Db 541 ATATTACCTTGGGTGTTACCAAGAGACCTCGCAGATCGACTTGTACAGTGCATGTGA 600  
QY 1065 TCCTGAGTGTGGTATCCAGTTTGTCAAAATACCTTGTATCCGTCACAACTTTGGCT 1124  
Db 601 CCTTGAGTGTGGTGTCTCAGTTTGTCAAAATACCTTGTATCCGTCACAACTTTGGCT 660  
QY 1125 GGAAGGGAATATGAAAGAACCAACCAAGAGCTTGGCTTCAAAATACCTTGTATGTGGAGT 1184  
Db 661 AGAAAGAAATATGAAAGAACCAACCAAGAGCTTGGCTTCAAAATACCTTGTATGTGGAGT 720  
QY 1185 CCATGTACAGCCACTGTACAAAGTGGGACAGAAAGAGCTTCCATCCATGAGGAATA 1244  
Db 721 CCATGTACAGCCACTGTACAAAGTGGGACAGAAAGAGCTTCCATCCATGAGGAATA 780  
QY 1245 CATGTACAGCTTGAAGAACATTTTTCAGCTTCTCGAAGCGAGAAATGAAAGTGAATAAA 1304  
Db 781 CATGTGCAATTTGAAGAACATTTTTCAGCTTCTCGAAGCGAGAAATGAAAGTGAATAAA 840  
QY 1305 AAGAGTGTATCTGGCCACTGTATGACCTTCTTTTAAAGGAGGCAAAAGTACTC 1364  
Db 841 AAGAGTGTATTTGGCCACAGATGACCTTCTTTTAAAGGAGGCAAAAGTACTC 900  
QY 1365 CAATTATGAATTTATTAGTATTAATCTATTTTCTTGCTAGCTGGAC-TACACAACCG 1421  
Db 901 CAATTATGAATTTAAATGAATCTATTTTCCCTGCTCAGCGGGACTTGGCCATCGN 960  
QY 1422 ATACACAGAAATTCACCTCGGGGG 1447  
Db 961 ATACCCCGAAATTCCTTCCGGGG 986

## RESULT 5

BQ653329  
LOCUS BQ653329 940 bp mRNA linear EST 15-JUL-2002  
DEFINITION AGENCOURT\_8298428 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:6270458  
5', mRNA sequence.

ACCESSION BQ653329

VERSION BQ653329.1 GI:21777501

KEYWORDS EST.

SOURCE Homo sapiens (human)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NIH-MGC <http://mgi.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)

Tissue Procurement: CGAP (Stanford)

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLM2445 row: k column: 03

High quality sequence stop: 700.

Location/Qualifiers

1..940

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:6270458"

/tissue type="Hepatocellular carcinoma, cell line"

/lab host="DH10B (phage-resistant)"

/clone lib="NIH MGC 100"

/note="Organ: liver; Vector: pOTB7; Site: 1; XhoI; Site 2:

EcoRI; cDNA made by oligo-dT priming. Directionally cloned

into EcoRI/XhoI sites using the following 5' adaptor:

GGCAGCAG(G). Size-selected >500bp for average insert size

1.8kb. Library constructed by Ling Hong in the laboratory

of Gerald M. Rubin (University of California, Berkeley)

using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."

BASE COUNT 286 a 197 c 222 g 235 t

## ORIGIN

Query Match 38.2%; Score 768; DB 13; Length 940;  
Best Local Similarity 92.0%; Pred. No. 5.6e-172;  
Matches 865; Conservative 0; Mismatches 70; Indels 5; Gaps 5;

QY 465 AGCTAAAGAGCTCTGGTTTTTCTACAAAGTGAATTAAGAAATTAAGAAATTAAGAGG 524

Db 1 AGCTAAAGAGCTCTGGTTTTTCTACAAAGTGAATTAAGAAATTAAGAAATTAAGAGG 60

QY 525 AAAAGAACTCCAAAGACATGACAGATGAATTTCTTTTGGATTAGACATCATGAAGGTC 584

Db 61 AAATGAATCTCAAAGACATGACAGATGAATTTCTTTTGGATTAGACATCATGAAGGTC 120

QY 585 TATCATGACAGATCTATATCTACCTCAGTCAAAAGATGAGCAGGTGCGCGGAAAA 644

Db 121 TATAATGAGGATCTATATCTACCTCAGTCAAAAGATGAGCAGGTGCGCGGAAAA 180

QY 645 AGAAGCCAAAGATCTGACAGAGCTGGTCCAGCGGAGAAATACATATCTGCAAGATCCAA 704

Db 181 AGAGGCCAAAGATCTGACAGAGCTGGTCCAGCGGAGAAATACATATCTGCAAGATCCAA 240

QY 705 GGACTGACAGCAAGACGAGAAAGCTGGTATTAATCAACAAAGCTGTGCTATGGATG 764

Db 241 GGACTGACAGCAAGACGAGAAAGCTGGTATTAATCAACAAAGCTGTGCTATGGATG 300

QY 765 TCAACTCCATCATGTGTCTTACTGCTTCATGATTTGCTTATGGCACCCAGCGAAACATCAT 824

Db 301 TCAGCTCCATCATGTGTCTTACTGCTTCATGATTTGCTTATGGCACCCAGCGAAACATCAT 360

QY 825 CTTGGAATCTCAGAAATTTGGCGCTATGCTACTGAGGAGTGGAGACTGTGTTAGACCTGT 884

Db 361 CTTGGAATCTCAGAAATTTGGCGCTATGCTACTGAGGAGTGGAGACTGTGTTAGACCTGT 420

QY 885 AAGTGACAGATGACAGACAGCTGGCTCTCCACTGGACACCTGGTCAGGTGAAGTGAA 944

Db 421 AAGTGACAGATGACAGACAGCTGGCTCTCCACTGGACACCTGGTCAGGTGAAGTGAA 480

QY 945 GGACAAAATATTTCAAGTGTGAGCTCCCATTTGTAGACAGCTTCCATCTCTGCTCTCC 1004

Db 481 GGACAAAATATTTCAAGTGTGAGCTCCCATTTGTAGACAGCTTCCATCTCTGCTCTCC 540

QY 1005 TTACTTACCTTGGCTGTACAGAAAGCTTGGAGATCGACTCTTGAGAGTCCATGTGA 1064

Db 541 ATATTTACCTTGGCTGTACAGAAAGCTTGGAGATCGACTCTTGAGAGTCCATGTGA 600

QY 1065 TCCTGCACTGTGGGTATCCAGTTTGTCAAAATCTTATGATCCGTCACAACTTTGGCT 1124

Db 601 CCCTGCACTGTGGGTATCCAGTTTGTCAAAATCTTATGATCCGTCACAACTTTGGCT 660

QY 1125 GGAAGGGAATATGAAAGAACCAACCAAGAGCTTGGCTTCAAAATCATCAGTTATTGGAGT 1184

Db 661 AGAAAGAAATATGAAAGAACCAACCAAGAGCTTGGCTTCAAAATCATCAGTTATTGGAGT 719

QY 1185 CCATGTACAGCGCACTGACAAAGTGGGAAACAGAGAGCTTCCATCCATGAGGAATA 1244

Db 720 CCATGTACAGCGCACTGACAAAGTGGGAAACAGAGAGCTTCCATCCATGAGGAATA 779

QY 1245 CATGGT-ACAGTTGAAGAAATTTTTCAGCTTCTCGAACCGAGATGAAGTGAATAA 1303

Db 780 CATGGTGAAGTGAAGAAATTTTTCAGCTTCTCGAACCGAGATGAAGTGAATAA 839

QY 1304 AAAGAGTGA-TCTGGCCACTGATGACCTTCTTTGT-TAAAGGAGGCAAAAGCAAGTA 1361

Db 840 AAAGAGTGA-TCTGGCCACTGATGACCTTCTTTGT-TAAAGGAGGCAAAAGCAAGTA 899

QY 1362 CTCCTCAATATGAA-TTATTAGTGAATCACTCTATTCTTTG 1400

Db 900 CCCCCTTATGAATTTATTATGATCACTCTATTCTTTG 939

```

RESULT 6
BX452321
LOCUS
DEFINITION
BX452321 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
clone CS0DGG05Y007 5-PRIME, mRNA sequence.
ACCESSION
BX452321
VERSION
BX452321.1 GI:31026371
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7499.f For
more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0BAG023ZH04_CS02144_1&cluster=7499.f.
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0BAG023ZH04_CS02144_1.
Location/Qualifiers
1..900
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DGG05Y007"
/tissue_type="B CELLS (RAMOS CELL LINE)"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT
250 a 201 c 200 g 245 t
ORIGIN
Query Match 37.7%; Score 757.6; DB 13; Length 900;
Best Local Similarity 91.6%; Pred. No. 1.7e-169;
Matches 810; Conservative 0; Mismatches 73; Indels 1; Gaps 1;

QY 742 AACAAAGGCTGGGTATGGATGTCACATCCATCATGTGGTTTACTGCTTCATGATTCT 801
Db |||||
15 AACAAAGGCTGGGTATGGCTGTGAGCTCCAT-ATGGTGCTACTGCTCATGATTGCA 73
QY |||||
802 TATGGCACCACCGAAGCACTCATCTTGAATCTCAGAAATGGCGCTACTGCTAGGGA 861
Db |||||
74 TATGGCACCACCGAAGCACTCATCTTGGATCTCAGATCTGGCTGTACTGGTGA 133
QY |||||
862 TGGGAGACTGTTTAGACTCTTAAGTAGAGATCGACAGAGCTCTGGCTCTCCACT 921
Db |||||
134 TGGGAGACTGTTATTTAGGCTGTAAGTAGAGATCGACAGAGATCTGGCATCTCCACT 193
QY |||||
922 GGACACTGTGAGGTGAAGTAGAGCAAAAATGTTCAAGTGGTCGAGCTCCCATCTGA 981
Db |||||
194 GGACACTGTGAGGTGAAGTAGAGCAAAAATGTTCAAGTGGTCGAGCTCCCATCTGA 253
QY |||||
982 GACAGCTCCATCTCGTCTCTTACTTACCTTGGCTGTACCAAGAGACCTTGCAGAT 1041
Db |||||
254 GACAGCTTTCATCCCGCTCTCCATATTACCTTGGCTGTACCAAGAGACCTTGCAGAT 313
QY |||||
1042 CGACTCCTGAGAGTCCATGGTGATCTCGAGTGGTGGTATCCAGTTTGTCAATATAC 1101
|||

```

```

Db 314 CGACTTGTACAGTGCATGGTGACCCCTGCAGTGTGGTGTCTCAGTTTGTCAATATAC 373
QY 1102 TTGATCCGTCCAAACCTTGGCTGGAAAGGGAATAGAGAAACCAACAAGAGCTTGGC 1161
Db |||||
374 TTGATCCGCCACACAGCTTGGCTAGAAAAGAAATAGAAAGACCAACAAGAGCTTGGC 433
QY 1162 TTCAAAATCCAGTATTGGAGTCCATGTAGAGCGCACTGACAAAGTGGGACAGAGCA 1221
Db |||||
434 TTCAAAATCCAGTATTGGAGTCCATGTAGAGCGCACTGACAAAGTGGGACAGAGCT 493
QY 1222 GCCTTCCATCCATTGGAGGAATACATGTCACCTTGAAGAAACATTTTTCAGCTTCTCGAA 1281
Db |||||
494 GCCTTCCATCCATTGGAGGAATACATGTCACCTTGAAGAAACATTTTTCAGCTTCTCGA 553
QY 1282 CGCAGAATGAAGTGGATAAAAAAGAGTGTATCTGGCCACTGTAGTACCCCTTCTTTGTTA 1341
Db |||||
554 CGCAGAATGAAGTGGATAAAAAAGAGTGTATCTGGCCACTGTAGTACCCCTTCTTTATTA 613
QY 1342 AAGGAGGCAAGAACAAAGTACTCCAAATATGAATTTATTAGTATTAATCTTATTTCTTGG 1401
Db |||||
614 AAGGAGGCAAGAACAAAGTACTCCCAATTTATTAAGTATTAATCTTATTTCTTGG 673
QY 1402 TCAGCTGGACTACACAAACCGATACACAGAAATTCACCTCGGGCGGTGATCCTCGATATA 1461
Db |||||
674 TCAGCTGGACTGCAATCGAATCGAATACACAGAAATTCACCTCGGGAGTATCCTCGATATA 733
QY 1462 CACTTCTCTCCCAAGGCTGACTTCTCTTGTGTGTAATTTTTCATCCAGGCTGTAGGGTT 1521
Db |||||
734 CATTTCTCTCAAGCAGACTTCTTAGTGTGTAATTTTTCATNCCCAGTCTGGCGAGTT 793
QY 1522 GCTTATGAATCATGCAAAACATCGCATCTGTAGTGCCTCTGCAAACTTCCATCTTTAGAT 1581
Db |||||
794 GCTTATGAATCATGCAAAACATCATCTGATGCCTCTGCAAACTTCCATCTTTAGAT 853
QY 1582 GACATCTACTATTTTGGAGGCGCAAAATGCCACAAACAGATTGC 1625
Db |||||
854 GACATTTACTATTTTGGGCGCGAATGCCCACAAATTC 897

RESULT 7
BX423860
LOCUS
DEFINITION
BX423860 886 bp mRNA linear EST 23-MAY-2002
AGENCOURT 7894821 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6158200
5', mRNA sequence.
ACCESSION
BX423860
VERSION
BX423860.1 GI:21119175
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS
NIH-MGC http://mgi.nci.nih.gov/
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC/DCTb/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13505 row: e column: 17
High quality sequence stop: 641.
Location/Qualifiers
1..886
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6158200"
/tissue_type="melanotic melanoma"
FEATURES
source

```

```
/lab host="DH10B (phage-resistant)"
/clone lib="NIH MGC 72"
/notes="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."
```

BASE COUNT 273 a 185 c 212 g 215 t 1 others  
ORIGIN

```
Query Match 37.4%; Score 751.2; DB 13; Length 886;
Best Local Similarity 91.9%; Pred. No. 5.6e-168;
Matches 804; Conservative 0; Mismatches 69; Indels 2; Gaps 1;

QY 507 ATTAAGAAATAGAAAGAAACGAACCTCCAAGACATCAGATGAAATCTTTTGGATT 566
Db 1 ATTAAGAACTTAGAAGGAATGAATCTCAAGACATCAGATGAAATCTTTTGGATT 60

QY 567 AGGACATCATGAAGGTCTATCATGACAGATCTATCTACTCTCAGTCAAAACAGATGGAGC 626
Db 61 AGGACATCATGAAGGTCTATAATGACGGATCTATACTACTCTCAGTCAAGATGGAGC 120

QY 627 AGGTGAGTGGCGGAAAGAAAGCCAAAGATCTGACAGCTGGTCCAGCGGAGATAAC 686
Db 121 AGGTGATTGGCGGAAAGAAAGCCAAAGATCTGACAGACTGGTTCAGCGGAGATAAC 180

QY 687 ATATCTGACGAATCCCAAGGACTGCGCAAAAGCCAGAAAGCTGGTATGTATATCAACA 746
Db 181 ATATCTTCAAGATCCCAAGGACTGCGCAAAAGCCAAAGCTGGTATGTATATCAACA 240

QY 747 AGGTGCTGGCTATGATGTCAATCCATCATGATGTTTACTGCTTCATGATGCTTATGG 806
Db 241 AGGTGCTGGCTATGCTGCTGCTCATCTCATGCTGCTTCTGCTTCATGATGCTATGG 300

QY 807 CACCAGCGAACACTCATCTTGGATCTCAGAAATTTGGCGCTATGCTACTGGAGGTGGA 866
Db 301 CACCAGCGAACACTCATCTTGGATCTCAGAAATTTGGCGCTATGCTACTGGAGGTGGA 360

QY 867 GACTGTGTTTAGACCTGTAAGTGACATGACACAGAGCTGCGCTCTCCACTGGACA 926
Db 361 GACTGTATTAGGCTGTATGTGACATGACACAGAGCTGCGCTCTCCACTGGACA 420

QY 927 CTGGTCAGTGAAGTGAAGGACAAAAATGTTCAAGTGTGCGAGTCCCAATGTAGACAG 986
Db 421 CTGGTCAGTGAAGTGAAGGACAAAAATGTTCAAGTGTGCGAGTCCCAATGTAGACAG 480

QY 987 CTTCCATCTCTCTCTCTTACTTACTCTTGGCTGTACAGAGACCTTGCGATCGACT 1046
Db 481 TCTTCATCCCTCTCTCTTACTTACTTGGCTGTACAGAGACCTTGCGATCGACT 540

QY 1047 CTTGAGAGTCAATGTGTCTCTGAGTGTGGTGTATCCCAAGTTTGTCAAAATCTTGAT 1106
Db 541 TGTACAGTGCATGTGTGACCTCTGAGTGTGGTGTCTGATTTGTCAAAATCTTGAT 600

QY 1107 CGGTCCACACCTTGGCTGGAAAGGAAATAGAGAAACCAACCAAGAGCTTGGCTTCAA 1166
Db 601 CGGCCCCACAGCTTGGCTAGAAAAAGAAATAGAGAGCCCAAGAGCTTGGCTTCAA 660

QY 1167 ACATCCAGTTATTGAGTCCATGTGACAGCGACTCAGAAAGTGGGAAACAGAGAGCCTT 1226
Db 661 ACATCCAGTTATTGAGTCCATGTGACAGCGCAAGAGTGGGAAACAGAGAGCCTT 720

QY 1227 CCATCCCATTTAGGAATACATGGTACATGTTCAAGAACATTTTCAGCTTCTCGAACGCGAG 1286
Db 721 CCATCCCATTTAGGAATACATGGTGTGATGTTGAAAGACATTTTCAGCTTCTCGAACGCGAG 780

QY 1287 AATGAAGTGGTAATAAAGAGTGTATCTGGCCACTGTAGACCC - TTTCTTTGTTAAAG 1344
Db 781 AATGAAGTGGTAATAAAGAGTGTATTTGGCCACTGTAGACCCCTTCTTTTAAAG 840

QY 1345 GAGGCAAGACAAAGTACTTCCAAATTTAT 1379
Db 841 GAGGCAAGACAAAGTACTTCCAAATTTAT 875
```

RESULT 8  
BX429921

LOCUS  
DEFINITION

ACCESSION  
VERSION

KEYWORDS  
SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France

Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 7499.f For

more information about this cluster, see

http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0BAA012ZA08\_CS01114\_1&cluster=7499.f.

Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Paradise Avenue Genoscope sequence ID : CS0BAA012ZA08\_CS01114\_1.

Location/Qualifiers

1..923

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0D005Y007"

/tissue\_type="B CELLS (RAMOS CELL LINE)"

/cell\_line="RAMOS CELL LINE"

/clone\_id="Homo sapiens B CELLS (RAMOS CELL LINE)"

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed

with a NotI-oligo(dT) primer. Five prime end enriched,

double-strand cDNA was digested with Not I and cloned into

the Not I and EcoRV sites of the pCMVSPORT 6 vector.

Library was not normalized."

BASE COUNT 285 a 184 c 231 g 218 t 5 others

ORIGIN

Query Match 36.9%; Score 741; DB 13; Length 923;

Best Local Similarity 92.5%; Pred. No. 1.5e-165;

Matches 821; Conservative 0; Mismatches 63; Indels 4; Gaps 4;

BX429921 923 bp mRNA linear EST 15-MAY-2003  
clone CS0D005Y007 5-PRIME, mRNA sequence.

BX429921  
EST.

Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France

Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 7499.f For

more information about this cluster, see

http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0BAA012ZA08\_CS01114\_1&cluster=7499.f.

Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Paradise Avenue Genoscope sequence ID : CS0BAA012ZA08\_CS01114\_1.

Location/Qualifiers

1..923

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0D005Y007"

/tissue\_type="B CELLS (RAMOS CELL LINE)"

/cell\_line="RAMOS CELL LINE"

/clone\_id="Homo sapiens B CELLS (RAMOS CELL LINE)"

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed

with a NotI-oligo(dT) primer. Five prime end enriched,

double-strand cDNA was digested with Not I and cloned into

the Not I and EcoRV sites of the pCMVSPORT 6 vector.

Library was not normalized."

BASE COUNT 285 a 184 c 231 g 218 t 5 others

ORIGIN

Query Match 36.9%; Score 741; DB 13; Length 923;

Best Local Similarity 92.5%; Pred. No. 1.5e-165;

Matches 821; Conservative 0; Mismatches 63; Indels 4; Gaps 4;

OY 423 GGGAAAGGATCATG-AAATCTTAAGGAGGAGATTGAAATGGAGCTTAAGAGCTCTGGT 481

Db 31 GGGNAAGNATCATGNAATCTCTGAGGAGGAGATTGAAATGGAGCTTAAGAGCTCTGGT 90

OY 482 TTTTCTTCAAAAGTGAATTTGAAGAAATTAAGAAATTTAGAAAGGAAACGAATCCAAAGAC 541

Db 91 TTTTCTTCAAGTGAATTTGAAGAAATTTAGAAATTTAGAAAGGAAATGAATCCAAAGAC 150

OY 542 ATGCAGATGAATTTCTTTTGGATTTAGACATCATGAAAGCTCTATCATGACAGATCTAT 601

Db 151 ATGCAGATGAATTTCTTTTGGATTTAGACATCATGAAAGCTCTATATGACGATCTAT 210

BX429921

EST.

Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France

Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 7499.f For

more information about this cluster, see

http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0BAA012ZA08\_CS01114\_1&cluster=7499.f.

Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Paradise Avenue Genoscope sequence ID : CS0BAA012ZA08\_CS01114\_1.

Location/Qualifiers

1..923

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0D005Y007"

/tissue\_type="B CELLS (RAMOS CELL LINE)"

/cell\_line="RAMOS CELL LINE"

/clone\_id="Homo sapiens B CELLS (RAMOS CELL LINE)"

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed

with a NotI-oligo(dT) primer. Five prime end enriched,

double-strand cDNA was digested with Not I and cloned into

the Not I and EcoRV sites of the pCMVSPORT 6 vector.

Library was not normalized."

BASE COUNT 285 a 184 c 231 g 218 t 5 others

ORIGIN

Query Match 36.9%; Score 741; DB 13; Length 923;

Best Local Similarity 92.5%; Pred. No. 1.5e-165;

Matches 821; Conservative 0; Mismatches 63; Indels 4; Gaps 4;





```
RESULT 10
BU184296
LOCUS
DEFINITION AGENCOURT_7971158 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6171012
5', mRNA sequence.
ACCESSION
VERSION BU184296
KEYWORDS
SOURCE EST.
ORGANISM Homo sapiens (human)
Homosapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 855)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL13538 row: k column: 13
High quality sequence stop: 711.
Location/Qualifiers
1..855
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6171012"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 67"
/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Library constructed by Life
Technologies."
BASE COUNT 265 a 172 c 224 g 193 t 1 others
ORIGIN
Query Match 36.8%; Score 738; DB 13; Length 855;
Best Local Similarity 92.2%; Pred. No. 7.7e-165;
Matches 777; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
QY 239 TTGAAAGCTGGAGCGCTTAAACACAAATGAGACTTGAGGAGAAATGGCTGAGTCTC 298
DB 12 TGGCAAGCTTGAAGCGTTTAAACACAAATGAGACTTGAGGAGAAATGGCGGAATG 71
QY 299 TCCGAATACCAAGAGCGCTTATGATCAGGGGACAGCTACAGGAGAGTCCGTTGTTAG 358
DB 72 TCCGGATACCAAGAGCGCTTATGATCAGGGGACAGCTATAGGAGAGTACGCGTTTATG 131
QY 359 AAGAACAGCTGTTAAGGCCAAGAACAGATTGAAATTTACAGAAACAGCTAGGAATG 418
DB 132 AAGAGCAGCTGTTAAGGCCAAGAACAGATTGAAATTTACAGAAACAGACCAAGAAATG 191
QY 419 ATCTGGGAAGGATCATGAATCTTTAAGGAGGAGGATTGAAATGGAGCTTAAGAGCTCT 478
DB 192 GTCTGGGAAGGATCATGAATCTTTAAGGAGGAGGATTGAAATGGAGCTTAAGAGCTCT 251
QY 479 GGTGTTTTCTCAAGAGTGAATTAAGAAATTAAGAAATTAAGAAACGAATCCCAA 538
DB 252 GGTGTTTTCTCAAGAGTGAATTAAGAAATTAAGAAATTAAGAAATTAAGCTCCAA 311
QY 539 GACATGCAGATGAATCTTTTGGATTAGGACATCATGAAGGTCTATCATGACAGATC 598
DB 312 GACATGCAGATGAATCTTTTGGATTAGGACATCATGAAGGTCTATATGACGGATC 371
599 TATACTACCTCAGTCAAAACAGATGGAGCAGGTGAGTCGGCGGAAAAAGAGCCAAAGATC 658
372 TATACTACCTCAGTCAACACAGATGGAGCAGGTGATTCGGCGGAAAAAGAGCCAAAGATC 431
659 TGACAGAGCTGGTCCAGCGGAGATACATATCTGCAGAGATCCCAAGGACTGCAGCAAG 718
432 TGACAGAGAACTGGTTCAGCGGAGATACATATCTTCAGAAATCCCAAGGACTGCAGCAAG 491
719 CCAGAAAGCTGGTATGTAATATCAACAAAGGCTGTGGCTATGGATGTCAACTCCATCATG 778
492 CCAGAAAGCTGGTGTGTAATATCAACAAAGGCTGTGGCTATGGCTGTGCTCATCATG 551
779 TGGTTTACTGCTTCATGATTCGTTATGGACCCAGCAACACTCATCTTGAATCTCAGA 838
552 TGGTCTACTGCTTCATGATTCGTTATGGACCCAGCAACACTCATCTTGAATCTCAGA 611
839 ATTGGCCCTATGCTACTGGAGGATGGAGACTGTGTTAGACCTGTAGTGAGACATGCA 898
612 ATTGGCCCTATGCTACTGGTGGATGGAGACTGTGTTAGGCTGTAGTGAGACATGCA 671
899 CAGACAGGTCTGGGCTCTCCACTGGACACTGGTCAGGTGAAGTGAAGGACAAAAATGTT 958
672 AAGACAGATCTGGCACTCTCCACTGGACACTGGTCAGGTGAAGTGAAGGACAAAAATGTT 731
959 AAGTGGTGGAGCTCCCACTTGTAGACAGCTCCATCTCGTCTCTTACTTACCTTGG 1018
732 AAGTGGTGGAGCTCCCACTTGTAGACAGCTCTTATCCCGTCTCTTATTTACCTTGG 791
1019 CTGTACCAAGAGCTTGCAGATCGACTCTCGAGAGTCCATCTGAGAGTCCATCTGAGTGTGGT 1078
792 CTGTACCAAGAGCTTGCAGATCGACTCTTGTACAGTGTGATGAGTGTGAGTGTGGT 851
1079 GGG 1081
852 GGG 854
RESULT 11
LOCUS
DEFINITION AGENCOURT_11259261 NIH_MGC_135 Mus musculus cDNA clone
IMAGE:30136221 5', mRNA sequence.
ACCESSION
VERSION CB194828
KEYWORDS EST.
SOURCE CB194828.1 GI:28220043
ORGANISM Mus musculus (house mouse)
REFERENCE
AUTHORS 1 (bases 1 to 898)
TITLE NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
cDNA Library Preparation: Invitrogen Corp
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM0035 row: h column: 22
High quality sequence stop: 658.
Location/Qualifiers
1..898
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30136221"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 135"
/note="Vector: pCMVSPORT6.1; Site 1: EcoRV; Site 2: NotI;
```



Normalized full-length enriched library from pooled mouse embryonic limb, maxilla and mandible, day 12.5, 13.5, 14.5, and 15.5 (size selected for the 0.5-1 kb fragments). Cloned directionally, priming method: Oligo-dr. cDNA enrichment: >1k bp. Average insert size 1.6k bp. Normalization (Cot value): 7.5 kb. Priming sequence: 5'-GACTATGCTTAGATCGGCGGCCCTTT-3'. Tissue contributed by David Rowe. Library constructed by ResGen, Invitrogen Corp."

```
BASE COUNT      272 a  203 c  193 g  230 t
ORIGIN
Query Match      36.4%; Score 730.2; DB 14; Length 898;
Best Local Similarity 93.4%; Pred. No. 5.6e-163;
Matches 784; Conservative 0; Mismatches 53; Indels 2; Gaps 2;

Qy 998 GTCCCTCTTACTTACCTTGGCTGTACAGAGACCTTGCAGATCGACTCTGAGAGTCC 1057
Db 1 GGCCTCTTACTTACCACTGGCTGTTCAGAGAGACCTTGCAGACGACTCTCTAAAGTCC 60

Qy 1058 ATGGTGATCTTCAGTGTGGTGGTATCCAGTTGTCAAAATACCTTGTATCGTCCACAA 1117
Db 61 ATGGTGATCTTCAGTGTGGTGGTATCCAGTTGTCAAAATACCTTGTATCGTCCACAA 120

Qy 1118 CTTGGCTGGAAGGGAATAGAGAAACCCACCAAGACTTGGCTTCAAAATACCAAGTGA 1177
Db 121 CTTGGCTGGAAGGGAATAGAGAAACCCACCAAGACTTGGCTTCAAAATACCAAGTGA 180

Qy 1178 TTGGAGTCACTGATGAGCACTGACAAAGTGGGAACAGAGACGCTTCCATCCATTTG 1237
Db 181 TTGGAGTCACTGATGAGCACTGACAAAGTGGGAACAGAGACGCTTCCATCCATTTG 240

Qy 1238 AGGAATACATGTACAGCTTGAAGAACATTTTCAGCTTCTGCAAGCAGAGATGAAGTGG 1297
Db 241 AGGAATACATGTACAGCTTGAAGAACATTTTCAGCTTCTGCAAGCAGAGATGAAGTGG 300

Qy 1298 ATAAAAAAGAGTGTATCTGGCCACTGATGACCCCTTCTTTGTTAAAGGAGGCAAGACAA 1357
Db 301 ATAAAAAAGAGTGTATCTGGCCACTGATGACCCCTTCTTTGTTAAAGGAGGCAAGACAA 360

Qy 1358 AGTACTCCAAATATGAATTTATAGTATACTCTATTTCTTGTGTCAGCTGGACTACACA 1417
Db 361 AGTACTCCAAATATGAATTTATAGTATACTCTATTTCTTGTGTCAGCTGGACTACACA 420

Qy 1418 ACCGATACACAGAAAATTCACCTTCGGGGCTGATCCTGATATACACTTTCTCTCCAGG 1477
Db 421 ATCGGTACACAGAAAATTCACCTTCGGGGCTGATCCTGATATACACTTTCTCTCCAGG 480

Qy 1478 CTGACTTCTTGTGTACTTTTTCATCCAGGCTGTAGGTTGCTTATGAAATCATGCG 1537
Db 481 CTGACTTCTTGTGTACTTTTTCATCCAGGCTGTAGGTTGCTTATGAAATCATGCG 540

Qy 1538 AAACHATGATCCTGATGCTCTGAAATTCATTTCTTTAGATGACATCTACTATTTTG 1597
Db 541 AAACHATGATCCTGATGCTCTGAAATTCATTTCTTTAGATGACATCTACTATTTTG 600

Qy 1598 GAGGCCAAAATGCCACACACAGATTGCGATTTTATCTTCCACCACTCGAATCAAGAGG 1657
Db 601 GAGGCCAAAATGCCACACACAGATTGCGATTTTATCTTCCACCACTCGAATCAAGAGG 660

Qy 1658 AAATCCCATGGAACCTGGAGATATCATTTGGTGTGGCTGGAACCATTTGGAATGGTTACT 1717
Db 661 AAATCCCATGGAACCTGGAGATATCATTTGGTGTGGCTGGAACCATTTGGAATGGTTACT 720

Qy 1718 CTAAGGTGTCAACAGAAAATCTAGAAAAAAGGCTGTACCCCTTCTTCAAAAGTCCGAG 1777
Db 721 CTAAGGTGTCAACAGAAAATCTAGAAAAAAGGCTGTATATCCCTTCTTCAAAAGTCCGAG 780

Qy 1778 AGAAGATAGAAACAG-TCAATATACCTTACATATCTG-AAGCTGAAAAATAGAGATGA 1834
Db 781 AGAAGATAGAAACAGTTCAGTATCCCATATCTCTGAAGCTGAAAAATAGAGATGA 839
```

## RESULT 12

```
AUI24128
LOCUS      AUI24128 NT2RM2 Homo sapiens cDNA clone NT2RM2001714 5', mRNA
DEFINITION AUI24128 NT2RM2 Homo sapiens cDNA clone NT2RM2001714 5', mRNA
sequence.
ACCESSION AUI24128
VERSION   AUI24128.1 GI:10948844
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
           Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (Bases 1 to 834)
AUTHORS   Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
           Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
           Isogai,T.
TITLE     HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
           Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
           Y., Sugano,S., Isogai,T.)
JOURNAL   Unpublished
COMMENT   Contact: Takao Isogai
           Genomics Laboratory
           Helix Research Institute
           1332-3 Yana, Kisarazu, Chiba 292-0812, Japan
           Tel: 81-438-52-3975
           Fax: 81-438-52-3986
           Email: genomics@hri.co.jp
           HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
           Research Institute; cDNA library construction; Department of
           Virology, Institute of Medical Science, University of Tokyo, and
           Helix Research Institute.
FEATURES   Location/Qualifiers
            1..834
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /clone="NT2RM2001714"
             /cell_type="teratocarcinoma"
             /cell_line="NT2"
             /clone_lib="NT2RM2"
             /note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
             precursor cells"
BASE COUNT 249 a 187 c 181 g 215 t
ORIGIN
Query Match      34.3%; Score 689.2; DB 9; Length 834;
Best Local Similarity 90.2%; Pred. No. 3.3e-153;
Matches 747; Conservative 0; Mismatches 80; Indels 1; Gaps 1;

Qy 1004 CTTACTTACCTTGGCTGTATCCAGAGACCTTGCAGATCGACTCCTGAGATCCATGGTG 1063
Db 1 CATATTTTACCTTGGCTGTATCCAGAGACCTTGCAGATCGACTTGTAGAGTGCATGGTG 60

Qy 1064 ATCTGCAAGTGTGGTGGTATCCAGTTTGTCAATACTTGTATCGATCCGTCACAACTTGGC 1123
Db 61 ACCCTGCAAGTGTGGTGGTGTCTCAGTTTGTCAATACTTGTATCGATCCGTCACAACTTGGC 120

Qy 1124 TGGAAAGGGAATAGAGAAACCAACCAAGAGCTTGGCTTCAAAATCTCCAGTATTGGAG 1183
Db 121 TAGAAAAAGAAATAGAGAAAGCCACCAAGAGCTTGGCTTCAAAATCTCCAGTATTGGAG 180

Qy 1184 TCCATGTGACAGCGCATGACAAAGTGGGAACAGAGCAGCCTTCCATCCCATTTGAGGAAT 1243
Db 181 TCCATGTGACAGCGCACAGACAAAGTGGGAACAGAGCAGCCTTCCATCCCATTTGAGGAAT 240

Qy 1244 ACATGGTACAGTTGAAGAACATTTTTCAGTTTCTCGAACCGAGATGAAGTGCATATAA 1303
Db 241 ACATGGTGCATTTGAAGAACATTTTTCAGTTTCTCGAACCGAGATGAAGTGCATATAA 300

Qy 1304 AAAGAGTGTATCTGGCCACTGATGACCTTCTTTGTTAAAGGAGGCAAGACAAAGTACT 1363
Db 301 AAAGAGTGTATTTGGCCACAGATGACCTTCTTTTAAAGGAGGCAAGACAAAGTACT 360
```



```

ACCESSION   BX414472
VERSION     BX414472.1  GI:30653877
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 757)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7499.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAN001ZF02_AN062_1&cluster=7499.f.
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0BAN001ZF02_AN062_1.
Location/Qualifiers
1. .757
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DG005Y007"
/tissue_type="B CELLS (RAMOS CELL LINE)"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT  191 a 160 c 167 g 238 t 1 others
ORIGIN
Query Match 31.6%; Score 634.6; DB 13; Length 757;
Best Local Similarity 90.1%; Pred. No. 3.3e-140;
Matches 679; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1098 ATACTTGATCCGTCACACACCTTGCTGGAAGGGAATAGAGAAACCAACCAAGACT 1157
DB 757 ATACTTGATCCGTCACACACCTTGCTGGAAGGGAATAGAGAAACCAACCAAGACT 698
QY 1158 TGGCTTCAAACTCCAGTTATTGGAGTCCATGTCAGACGCACTGCACAAAGTGGGAACAGA 1217
DB 697 TGGCTTCAAACTCCAGTTATTGGAGTCCATGTCAGACGCACTGCACAAAGTGGGAACAGA 638
QY 1218 AGCAGCTTCCATCCATGAGGATACATGTCACGTTGAGAACATTTTCAGCTTCT 1277
DB 637 AGCTGCTTCCATCCATGAGGATACATGTCGTCATGTTGAGAACATTTTCAGCTTCT 578
QY 1278 CGAAGCGAGAAAGTGGATAAAGAGAGTGTATCTGGCCACTGATGACCCCTTCTTT 1337
DB 577 TCACGCGAGATGCAAGTGGACAAAGAGAGTGTATTGGCCACAGATGACCCCTTCTTT 518
QY 1338 GTTAAAGGAGGCAAGACAAAGTACTCCAAATPATGAATTTATTAGTGATAACTCTATTTC 1397
DB 517 ATTAAAGGAGGCAAGACAAAGTACTCCCAATPATGAATTTATTAGTGATAACTCTATTTC 458
QY 1398 TTGGTCAGCTGCACTACACACCGATACACAGAAATTCACCTCGGCGGTGATCCCTGGA 1457
DB 457 CTGGTCAGCTGCACTGCACTACACCGATACACAGAAATTCACCTCGGCGGTGATCCCTGGA 398
QY 1458 TATACATTTCTCTCCAGGCTGACTTCTCTGTGTACTTTTTCATCCAGGCTCTGTAG 1517
DB 397 TATACATTTCTCTCTCAGGAGACTTCTCTGTGTACTTTTTCATCCAGGCTCTGTAG 338
QY 1518 GGTTCGCTTATGAATCATGCACAACTGCATCCTGATGCTCTGCACAACTTCCATTCTTT 1577

Db 337 AGTTGCTTATGAATTATGAACACTACATCCTGTATGCTCTCTCAAACTTCCATTCTTT 278
QY 1578 AGATGACATCTACTATTATTTGGAGGCCAAATATGCCACAAACAGATTTGCAGTTTATCTCTCA 1637
Db 277 AGATGACATCTACTATTATTTGGGGGCCAAGATGCCACAAATCAATTTGCCATTATGCTCA 218
QY 1638 CCAACCTCGAACTAAAGAGGAAATCCCATGGAACCTCGGAGATATCATTTGGTGTGGCTGG 1697
Db 217 CCAACCCCGAACTCGACATGAATTTCCCATGGAACCTCGGAGATATCATTTGGTGTGGCTGG 158
QY 1698 AAACCATGGAATGGTTACTCTAAAGCTGTCAACAGAAACTAGGAAACACAGCCCTGTA 1757
Db 157 AAATCATTTGGATGGCTATTCTTAAAGGTGTCAACAGGAAATTTGGGAAGGACGGGCTATA 98
QY 1758 CCCCTCTACAAAGCTCGAGAGAGATAGAAACAGTCAAAATACCTCATATCTCTGAAGC 1817
Db 97 TCCCTCTACAAAGTTCGAGAGAGATAGAAACAGTCAAGTCAAGTACCCACATATCTCTGAGC 38
QY 1818 TGAATAATAGAGATGGAGTGTAGAGATTAACAA 1851
Db 37 TGAGAAATAAGCTCAGATCGAGAGAGATAAACGA 4

RESULT 15
CB989769 790 bp mRNA linear EST 01-MAY-2003
LOCUS      AGENCOURT 13892856 NIH MGC 147 Homo sapiens cDNA clone
IMAGE:30344717 5', mRNA sequence.
ACCESSION  CB989769
VERSION     CB989769
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 790)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: csapbs-r@mail.nih.gov
Tissue Procurement: Dr. Stefan Hansson
cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help
and advice from Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM381 row: h column: 06
High quality sequence stop: 619.
Location/Qualifiers
1. .790
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30344717"
/tissue_type="Human Placenta"
/lab_host="DH10B Tona"
/clone_lib="NIH_MGC_147"
/notes="Organ: placenta; Vector: pBluescriptR; Site_1:
all-XhoI; Site_2: BamH; Oligo-dT primed using primer
5'-TTTTTTTTTTTTTTVN-3', size-selected for average
insert size 2.3 kb and normalized to 5. This is a
primary library enriched for full-length clones and
constructed using the Cap-trapper method (Carninci, in
preparation). Library constructed by M. Brownstein
(NIMH/NHGRI, National Institutes of Health). Note: This is
a NIH MGC library."
BASE COUNT  244 a 157 c 202 g 186 t 1 others
ORIGIN

```

Query Match		31.6%; Score 634.6; DB 14; Length 790;			
Best Local Similarity		93.0%; Pred. No. 3.3e-140;			
Matches 675; Conservative		0; Mismatches 50; Indels 1; Gaps 1;			
QY	357	AGAAGAACAGCTTGTTAAGGCCCAAGACAGATTGAAAATTACAAGAAACAAGCTAGGAA	416		
DB	27	AGNAGAGCAGCTTGTTAAGGCCCAAGACAGATTGAAAATTACAAGNACAGACCCAGAA	86		
QY	417	TGATCTGGAAAGGATCATGAAATCTTTAAGGAGGAGATTGAAAATGGAGCTAAAGAGCT	476		
DB	87	TGGTCTGGGAAAGGATCATGAAATCTCTGAGGAGGAGATTGAAAATGGAGCTAAAGAGCT	146		
QY	477	CTGGTTTTTCTACAAAGTGAATTGAAGAAATTAAGAAATTAAGAGGAAACCAACTCCA	536		
DB	147	CTGGTTTTTCTACAGAGTGAATTGAAGAAATTAAGAAATTAAGAGGAAATGACTCCA	206		
QY	537	AAGACATGCAGATGAAATCTTTTGGATTTAGGACATCATGAAAGGCTTATCATGACAGA	596		
DB	207	AAGACATGCAGATGAAATCTTTTGGATTTAGGACATCATGAAAGGCTTATAATGACGGA	266		
QY	597	TCTATACTACTCAGTCAAAACAGATGGAGCAGGTGAGTGGCGGGAAGAAAGACCCAAAGA	656		
DB	267	TCTATACTACTCAGTCAAGACAGATGGAGCAGGTGATTGGCGGGAAGAAAGAGGCCAAAGA	326		
QY	657	TCTGACAGAGCTGGTCCAGCGGAGAAATACATATCTGCAGAAATCCCAAGGACTGCAGCAA	716		
DB	327	TCTGACAGAACTGGTTCCGCGGAGAAATACATATCTTCAGAAATCCCAAGGACTGCAGCAA	386		
QY	717	AGCCAGAAAGCTGGTATGTAATATCAACAAAGGCTGTGGCTATGGATGTCAACTCCATCA	776		
DB	387	AGCCAAAAGCTGGTGTGTAATATCAACAAAGGCTGTGGCTATGGCTGTCAAGTCCATCA	446		
QY	777	TGTGGTTTACTGCTTCATGATTGCTTATGGCACCCAGCGAACACTCATCTTGAATCTCA	836		
DB	447	TGTGGTCTACTGCTTCATGATTGCTATGGCACCCAGCGAACACTCATCTTGAATCTCA	506		
QY	837	GAATTGGCGCTATGCTACTGGAGGATGGGAGACTGTCTTTAGACCTGTAAGTGAGACATG	896		
DB	507	GAATTGGCGCTATGCTACTGGTGGATGGGAGACTGTATTTAGGCCCTGTAAGTGAGACATG	566		
QY	897	CACAGACAGGTCTGGCCCTCTCCACTGGACACTGGTCAGGTGAAGTGAAGGACAAAATGT	956		
DB	567	CACAGACAGATCTGGCATCTCCACTGGACACTGGTCAGGTGAAGTGAAGGACAAAATGT	626		
QY	957	TCAAGTGTTCAGCTCCCATTTGTAGACAGCTCCATCTCGTCCCTTACTTACCCCTT	1016		
DB	627	TCAAGTGGTTCAGCTCCCATTTGTAGACAGTCTTCATCCCGCTCTCTCCACATTTACCCCTT	686		
QY	1017	GGCTGTACCCAGAAAGCTTGCAGATCGACTCTCTGCAGAGTCCAT-GGTGATCTCTGCAGTGT	1075		
DB	687	GGCTGTACCCAGAAAGCTTGCAGATCGACTTGTACAGTGTGATGGGTGACCCCTGCAGTGN	746		
QY	1076	GGTGGG 1081			
DB	747	TGGTGG 752			

Search completed: February 2, 2004, 18:36:52  
Job time : 4338.69 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 2, 2004, 10:19:28 ; Search time 6342.17 Seconds  
(without alignments)  
11146.322 Million cell updates/sec

Title: US-09-971-773-2  
Perfect score: 1728  
Sequence: 1 atggggcagtgactgttc.....atctgaagctgaaaaatag 1728

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.on.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1728	100.0	1728	6	BD168517	BD168517 Cells pro
2	1728	100.0	2976	10	BC010666	BC010666 Mus muscu
3	1723.2	99.7	2176	10	AB025198	AB025198 Mus muscu
4	1584	91.7	2008	6	BD168516	BD168516 Cells pro
5	1501.4	86.9	1759	6	AX662735	AX662735 Sequence
6	1501.4	86.9	1759	6	BD087745	BD087745 Plant cel
7	1501.4	86.9	2002	9	D89289	D89289 Homo sapien
8	1501.4	86.9	2100	6	E15725	E15725 Human mRNA
9	1501.4	86.9	2796	9	HSAS39535	AJ539535 Homo sapi
10	1501.4	86.9	2898	9	HSY17976	Y17976 Homo sapien
11	1501.4	86.9	2902	9	HSAS39536	AJ539536 Homo sapi
12	1501.4	86.9	2992	9	HSY17977	Y17977 Homo sapien
13	1501.4	86.9	3186	9	HSY17978	Y17978 Homo sapien
14	1501.4	86.9	3280	9	HSY17979	Y17979 Homo sapien
15	1501.4	86.9	3568	9	HSAS36055	AJ536055 Homo sapi
16	1501.4	86.9	3666	9	HSAS36053	AJ536053 Homo sapi
17	1501.4	86.9	3772	9	HSAS36054	AJ536054 Homo sapi
18	1501.4	86.9	4196	9	HSAS36056	AJ536056 Homo sapi
19	1496.6	86.6	2100	6	AR170077	AR170077 Sequence
20	1472.6	85.2	1728	4	AF247186	AF247186 Bos tauru
21	1469.4	85.0	1728	6	AR170074	AR170074 Sequence
22	1469.4	85.0	1728	6	E14720	E14720 Porcine mRN
23	1469.4	85.0	1836	4	D86723	D86723 Porcine mRN
24	1208	69.9	2682	9	BC025385	BC025385 Homo sapi
25	1088.2	63.0	2129	5	XLA514872	AJ514872 Xenopus l
26	941	54.5	1850	9	AF052088	AF052088 Homo sapi
27	931	53.9	979	6	BD168522	BD168522 Cells pro
28	894.2	51.7	979	6	BD168521	BD168521 Cells pro
29	876.2	50.7	1017	6	E43925	E43925 Antibody re
30	755.4	43.7	1772	9	AB049740	AB049740 Homo sapi
31	604.6	35.0	699	6	E43926	E43926 Antibody re
32	401.6	23.2	1262	9	HSAS14325	AJ514325 Homo sapi
33	401.6	23.2	1396	9	HSAS14324	AJ514324 Homo sapi
34	378.8	21.9	1860	3	AF411264	AF411264 Drosophil
35	378.8	21.9	2764	3	AY051451	AY051451 Drosophil
36	357.2	20.7	2860	3	AK113018	AK113018 Ciona int
37	322.6	18.7	844	11	BV033231	BV033231 S212P6677
38	322.6	18.7	171073	2	AC132342	AC132342 Mus muscu
39	316.6	18.3	110000	2	AC129041.1	Continuation (2 of
40	316.6	18.3	110000	2	AC129041.2	Continuation (3 of
41	316.6	18.3	150709	2	AC134206	AC134206 Rattus no
42	316.6	18.3	237954	2	AC128532	AC128532 Rattus no
43	296.4	17.2	394	6	BD071198	BD071198 Secreted
44	271	15.7	1147	9	AB049828	AB049828 Homo sapi
45	271	15.7	186158	9	CNS018P0	AL109847 Human chr

# ALIGNMENTS

RESULT 1  
BD168517  
LOCUS BD168517 1728 bp DNA linear PAT 17-JAN-2003  
DEFINITION Cells producing antibody composition.  
ACCESSION BD168517  
VERSION BD168517.1 GI:27874329  
KEYWORDS WO 0231140-A/2.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1728)  
AUTHORS Kanda, Y., Sato, M., Nakamura, K., Uchida, K., Shinkawa, T., Yamane, N.,  
Hosaka, E., Yamasaki, M. and Hanai, N.  
TITLE Cells producing antibody composition

JOURNAL	Patent: WO 0231140-A 2 18-APR-2002;	
COMMENT	OS KYOWA HAKKO KOGYO CO LTD	
	PN WO 0231140-A/2	
	PD 18-APR-2002	
	PF 05-OCT-2001 WO 2001JP008804	
	PR 06-OCT-2000 JP 00P 308526	
	PI YUTAKA KANDA, MITSUO SATO, KAZUYASU NAKAMURA, KAZUHISA UCHIDA, PI	
	TOYOHIDE SHINKAWA, NAOKO YAMANE, EMI HOSAKA, MOTOO YAMASAKI, NOBUO	
	PI HANA1	
	PC C12N5/10, C12P21/08, C07K16/00, A01K67/00, A61K39/395, C12N9/00, PC	
	C12N15/52,	
	PC GOIN33/53	
	CC Cells producing antibody composition	
	PH Key Location/Qualifiers	
	FT source 1..1728	
FEATURES	Location/Qualifiers	
source	1..1728	
	/organism='Mus musculus (mouse)'	
	/mol_type='genomic DNA'	
	/db_xref='taxon:10090'	
BASE COUNT	529 a 365 c 400 g 434 t	
ORIGIN		
	Query Match 100.0%; Score 1728; DB 6; Length 1728;	
	Best Local Similarity 100.0%; Pred. No. 0;	
	Matches 1728; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 ATGCGGGGATGACGTGGTTCCTGGGGTGGGATTATGCTCATTTTTCCTGGGGGACC 60	661 GGTGTCAACTCCATCAGCTGGTCTACTGTTTTCATGATTGCTTATGGCAACCAGCAACA 720
DB	1 ATGCGGGGATGACGTGGTTCCTGGGGTGGGATTATGCTCATTTTTCCTGGGGGACC 60	661 GGTGTCAACTCCATCAGCTGGTCTACTGTTTTCATGATTGCTTATGGCAACCAGCAACA 720
QY	61 TTGTTATTTTATATAGTGGTTCATTTGGTTCAGATATATGACACCCCTGATCAGTCCAGC 120	721 CTCATCTTTGGAATCTCAGAAATGGCGCTATGCTACTGTTGGATGGAGACTGTGTTTAGA 780
DB	61 TTGTTATTTTATATAGTGGTTCATTTGGTTCAGATATATGACACCCCTGATCAGTCCAGC 120	721 CTCATCTTTGGAATCTCAGAAATGGCGCTATGCTACTGTTGGATGGAGACTGTGTTTAGA 780
QY	121 AGAGAACTCTCCAAGATTCTTCAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 180	781 CCTGTAAGTGAGACATGTACAGACAGATCTGGCCCTCTCCACTGGACACTGTGTCAGTGAA 840
DB	121 AGAGAACTCTCCAAGATTCTTCAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 180	781 CCTGTAAGTGAGACATGTACAGACAGATCTGGCCCTCTCCACTGGACACTGTGTCAGTGAA 840
QY	181 AGCGAATGGCTGAGTCTCTCGAATACCAAGAGCCCATGACAGGGGACAGCTACA 240	841 GTAATGACAAAACATTTCAAGTGGTCCAGCTCCCCATTTAGACAGAGCTCCATCCTCGG 900
DB	181 AGCGAATGGCTGAGTCTCTCGAATACCAAGAGCCCATGACAGGGGACAGCTACA 240	841 GTAATGACAAAACATTTCAAGTGGTCCAGCTCCCCATTTAGACAGAGCTCCATCCTCGG 900
QY	241 GGAAGAGTCCGTTTGAAGAACAGCTTGTAAAGGCCAAAGAACAGATTCGAAATTTAC 300	901 CCTCCTTACTTACCACTGGCTGTTCCAGAGACCTTGCAGACCGACTCTTAAGAGTCCAT 960
DB	241 GGAAGAGTCCGTTTGAAGAACAGCTTGTAAAGGCCAAAGAACAGATTCGAAATTTAC 300	901 CCTCCTTACTTACCACTGGCTGTTCCAGAGACCTTGCAGACCGACTCTTAAGAGTCCAT 960
QY	301 AAGAACAAGCTAGAAATGGTCTGGGAGAGATCATGAAATCTTAAAGAGGAGATTGAA 360	961 GGTGACCCCTGAGTGGTGGTGTCCAGATTTGTCAAAATACTTGTGATTCGTCACAACCT 1020
DB	301 AAGAACAAGCTAGAAATGGTCTGGGAGAGATCATGAAATCTTAAAGAGGAGATTGAA 360	961 GGTGACCCCTGAGTGGTGGTGTCCAGATTTGTCAAAATACTTGTGATTCGTCACAACCT 1020
QY	361 AATGAGCTAAAGAGCTGTGTTTCTCAAAGCGAATCTGAAAGAAATTAAGCATTTA 420	1021 TGGCTGGAAGAGGAAATAGAAAGACCCACCAAGAGCTTGGCTTCAAAATCAGTCAATTT 1080
DB	361 AATGAGCTAAAGAGCTGTGTTTCTCAAAGCGAATCTGAAAGAAATTAAGCATTTA 420	1021 TGGCTGGAAGAGGAAATAGAAAGACCCACCAAGAGCTTGGCTTCAAAATCAGTCAATTT 1080
QY	421 GAAGGAATGAACTCCAAAGACATCGAGATGAAATCTTTTGGATTTAGGACACCATGAA 480	1081 GGAGTCCATGTCCAGCGCACAGCAAAAGTGGGAAACAGAAAGCCTTCCACCCATCGAG 1140
DB	421 GAAGGAATGAACTCCAAAGACATCGAGATGAAATCTTTTGGATTTAGGACACCATGAA 480	1081 GGAGTCCATGTCCAGCGCACAGCAAAAGTGGGAAACAGAAAGCCTTCCACCCATCGAG 1140
QY	481 AGGTCTATCATGACAGATCTACTACCTCAGTCAAAACAGATGGAGCGGATTTGGCGT 540	1141 GAGTACATGGTACAGCTTGAAGAACATTTTCAGCTTCTCGCACGAGAAATGCAAGTGGAT 1200
DB	481 AGGTCTATCATGACAGATCTACTACCTCAGTCAAAACAGATGGAGCGGATTTGGCGT 540	1141 GAGTACATGGTACAGCTTGAAGAACATTTTCAGCTTCTCGCACGAGAAATGCAAGTGGAT 1200
QY	541 GAAAAGAGGCCAAAGATCTGAAGAGCTGGTCCAGCGGAGAAATACATATCTCCAGAT 600	1201 AAAAAAGAGTATATCTGGCTACTGATCTTCTTGTAAAGGAGGCAAGACAAAG 1260
DB	541 GAAAAGAGGCCAAAGATCTGAAGAGCTGGTCCAGCGGAGAAATACATATCTCCAGAT 600	1201 AAAAAAGAGTATATCTGGCTACTGATCTTCTTGTAAAGGAGGCAAGACAAAG 1260
QY	601 CCTTAGGAGTCAGCAAGCCAGGAGCTGGTGTGTAACATCAATAAAGGCTGTGCTAT 660	1261 TACTCCATTAATGAATTTATTAGTGATAACTCTATTTCTGTCAGCTGGACTACACAAT 1320
DB	601 CCTTAGGAGTCAGCAAGCCAGGAGCTGGTGTGTAACATCAATAAAGGCTGTGCTAT 660	1261 TACTCCATTAATGAATTTATTAGTGATAACTCTATTTCTGTCAGCTGGACTACACAAT 1320
		1321 CGGTACACAGAAAAATTCATCTCGGGGTGTGATCTCTGGATATACACTTTCTCTCAGAGCT 1380
		1321 CGGTACACAGAAAAATTCATCTCGGGGTGTGATCTCTGGATATACACTTTCTCTCAGAGCT 1380
		1381 GACTTTCTAGTGTGTAATTTTTCATCCAGGTCTGTGCGGGTGTCTTATGAAATCATGCA 1440
		1381 GACTTTCTAGTGTGTAATTTTTCATCCAGGTCTGTGCGGGTGTCTTATGAAATCATGCA 1440
		1441 ACCCTGCATCTGATGCTCTCGCAATCTTCCATCTTTTGGATGACATCTACTATTTTGA 1500
		1441 ACCCTGCATCTGATGCTCTCGCAATCTTCCATCTTTTGGATGACATCTACTATTTTGA 1500
		1501 GGCCAAAATGCCCAATCAGATTGCTGTTTATCTCAAAAACCTCGAACTGAAGAGGAA 1560
		1501 GGCCAAAATGCCCAATCAGATTGCTGTTTATCTCAAAAACCTCGAACTGAAGAGGAA 1560
		1561 ATTCCAATGGAACTGGAGATATCATTTGGTGTGGTGGGAAACCAATGGGATGGTTATCT 1620
		1561 ATTCCAATGGAACTGGAGATATCATTTGGTGTGGTGGGAAACCAATGGGATGGTTATCT 1620
		1621 AAGGTATCAACAGAAACTTGGAAACAGGCTTATATCCCTCTCAAAAGTCCGAGAG 1680
		1621 AAGGTATCAACAGAAACTTGGAAACAGGCTTATATCCCTCTCAAAAGTCCGAGAG 1680
		1681 AAGATAGAAACAGTCAAGTATCCCAATATCTCTGAAGCTGAAATAG 1728
		1681 AAGATAGAAACAGTCAAGTATCCCAATATCTCTGAAGCTGAAATAG 1728

## RESULT 2

BC010666 2976 bp mRNA linear ROD 16-APR-2003  
Mus musculus fucosyltransferase 8, mRNA (cDNA clone MGC:11418  
IMAGE:3594582), complete cds.  
ACCESSION BC010666  
VERSION BC010666.1 GI:14715012  
KEYWORDS MGC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

## REFERENCE

1 (bases 1 to 2976)  
Klausner, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Sheets, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Sahegy, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,  
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E.,  
Gnerich, A., Schein, J.E., Jones, S.J. and Marra, M.A.

## AUTHORS

Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
22388257  
12477932  
2 (bases 1 to 2976)  
Strausberg, R.  
Direct Submission  
Submitted (10-JUL-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [c9apbs-remail.nih.gov](mailto:c9apbs-remail.nih.gov)  
Tissue Procurement: Jeffrey Green M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunsberg, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulsegged, H.,  
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
A.N., Gibbs, R.A.

## TITLE

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LILNL at: <http://image.llnl.gov>  
Series: IRAP Plate: 16 Row: 1 Column: 5  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 8393370.  
Location/Qualifiers  
1. .2976  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="MGC:11418 IMAGE:3594582"  
/tissue\_type="Mammary tumor, C3(1)-Tag model. Infiltrating  
ductal carcinoma. 5 month old virgin mouse."  
/clone\_lib="NCI\_CGAP\_Mam6"

## REMARK

COMMENT

## FEATURES

source

## gene

/lab host="DH10B"  
/note="Vector: pCMV-SPORT6"  
1. .2976  
/gene="Fut8"  
/db\_xref="LocusID:53618"  
/db\_xref="MGI:1858901"  
584. .2311  
/codon\_start=1  
/product="fucosyltransferase 8"  
/protein\_id="AAH10666.1"  
/db\_xref="GI:14715013"  
/db\_xref="LocusID:53618"  
/translation="MRATGSGRWIMLILFAWGTLLFYIGHVLVDNDHPHSSRELS  
KIALKGLKQONEDLRMAESLRIPGPIDQGTATGRVRVLEBQLVAKQIENYKK  
QRANGLGDKDEHLLRRRTENGAKELWFFLOSELKKLKHLEGNLOHDAEILLDLGHHE  
RSIMDLVYLSOTDAGDWEKADJTELVRRTVLYONPKDCKARKLVNKNKG  
GYGQOLHIVYCFMIAVTQRTLLLESONWRYATGGWETVFRVSETCTDRSGLSTGH  
WSGVNDKNIQVVELPVLSDVLSHPPPLPLAVPEDLARLLRVLHGDPAWVWSQVVKY  
LIRPQWLEKEIEEATKLGPKFVIGVHVRRTDKVGEAAAFHFIHRYMVHEEFLQ  
LARMQVDKRVLATDDPTLLKEAKTKYSNYEFTSDNISWSAGLHNRVTEENSLFGV  
ILDTHFLQADFLVCTFSSOVRVAYEIMOTLHPDASANFSLDDIYVFGQNAHNOI  
AVYPHKPTBEEIPEWEDDIIGVAGNHWGYSKGINRKLKGLTGLYPSYKVRKEITFVK  
YPTYPEAKK"  
BASE COUNT 909 a 626 c 660 g 781 t  
ORIGIN  
Query Match 100.0%; Score 1728; DB 10; Length 2976;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1728; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGCGGGCATGGACTGGTTCCTGGCGTTGGATTATGCTCATTTCTTTTGGCTGGGGACC 60  
DB 584 ATGCGGGCATGGACTGGTTCCTGGCGTTGGATTATGCTCATTTCTTTTGGCTGGGGACC 643  
QY 61 TTGTTATTTATATAGTGGTTCATTTGGTTCGAGATATGACCCCTGATCCTCCAGC 120  
DB 644 TTGTTATTTATATAGTGGTTCATTTGGTTCGAGATATGACCCCTGATCCTCCAGC 703  
QY 121 AGAGAACTCTCCAGATTCTTGCAAGCTTCAACCGCTTAAACAGCAAGAAATGAAGACTTG 180  
DB 704 AGAGAACTCTCCAGATTCTTGCAAGCTTCAACCGCTTAAACAGCAAGAAATGAAGACTTG 763  
QY 181 AGGCGAATGGCTGAGTCTCTCCGAATACCAAGGCCCCCATTTGACCGGGGACAGCTACA 240  
DB 764 AGGCGAATGGCTGAGTCTCTCCGAATACCAAGGCCCCCATTTGACCGGGGACAGCTACA 823  
QY 241 GGAAGAGTCCGTGTTTATAGAAAGACAGCTTCTTAAAGGCAAGAACAGATTTGAAAATTTAC 300  
DB 824 GGAAGAGTCCGTGTTTATAGAAAGACAGCTTCTTAAAGGCAAGAACAGATTTGAAAATTTAC 883  
QY 301 AAGAAACAGCTAGAAATGGTCTGGGGAAGGATCATGAAATCTTAAAGAGGAGGATTGAA 360  
DB 884 AAGAAACAGCTAGAAATGGTCTGGGGAAGGATCATGAAATCTTAAAGAGGAGGATTGAA 943  
QY 361 AATGAGGCTAAGAGCTCTGCTTTTCTCAAAAGCGAACTCAAGAAATTTAAAGCATTTA 420  
DB 944 AATGAGGCTAAGAGCTCTGCTTTTCTCAAAAGCGAACTCAAGAAATTTAAAGCATTTA 1003  
QY 421 GAAGAAATGAATCTTCAAGACATCGAGATGAAATCTTTTGGATTTAGGACCATGAA 480  
DB 1004 GAAGAAATGAATCTTCAAGACATCGAGATGAAATCTTTTGGATTTAGGACCATGAA 1063  
QY 481 AGGTCTATCATGACAGATCTATCTACTCTAGTCAACAGATGGAGCGAGGATTTGGCGT 540  
DB 1064 AGGTCTATCATGACAGATCTATCTACTCTAGTCAACAGATGGAGCGAGGATTTGGCGT 1123  
QY 541 GAAAAAGAGGCCAAGATCTCACAGAGCTGGTTCAGCGGAGAAATAACATATCTCCAGAA 600  
DB 1124 GAAAAAGAGGCCAAGATCTCACAGAGCTGGTTCAGCGGAGAAATAACATATCTCCAGAA 1183  
QY 601 CCTAAGGACTGCAGCAAGAGCCAGGAAGCTGGTGTGTATCAATCAATAAGGCTGTGGCTAT 660  
DB 1184 CCTAAGGACTGCAGCAAGAGCCAGGAAGCTGGTGTGTATCAATCAATAAGGCTGTGGCTAT 1243





241 QY GGAAGAGTCCGTGTTTAAAGAAACAGCTTGTAAAGGCCAAAGAACAGATTGAAATTTAC 300  
445 Db GGAAGAGTCCGTGTTTAAAGAAACAGCTTGTAAAGGCCAAAGAACAGATTGAAATTTAC 504  
301 QY AAGAAACAGCTAGAAATGCTGCGGGAAGATCATGAAATCTTAAAGAGAGGATTGAA 360  
505 Db AAGAAACAGCTAGAAATGCTGCGGGAAGATCATGAAATCTTAAAGAGAGGATTGAA 564  
361 QY AATGAGCTAAGAGCTGCTGTTTCTCAAGAGCGAATCTGAAAGAAATTAAGCATTTA 420  
565 Db AATGAGCTAAGAGCTGCTGTTTCTCAAGAGCGAATCTGAAAGAAATTAAGCATTTA 624  
421 QY GAAGAAATGAATCCAAAGACATGCAGATGAAATCTTTTGGATTTAGGACACCATGAA 480  
625 Db GAAGAAATGAATCCAAAGACATGCAGATGAAATCTTTTGGATTTAGGACACCATGAA 684  
481 QY AGGTCTATCATGACAGATCTATCTACCTCAGTCAACAGATGGAGCAGGGGATGGCGT 540  
685 Db AGGTCTATCATGACAGATCTATCTACCTCAGTCAACAGATGGAGCAGGGGATGGCGT 744  
541 QY GAAAGAGGCGCAAGATCTGACAGAGCTGCTCCAGCGGAGAAATCAATATCTCCAGAA 600  
745 Db GAAAGAGGCGCAAGATCTGACAGAGCTGCTCCAGCGGAGAAATCAATATCTCCAGAA 804  
601 QY CCTAAGGACTGCAGCAAGCCAGGAAGCTGTGTGTAAATCAATCAATGAAGGCTGTGGCTAT 660  
805 Db CCTAAGGACTGCAGCAAGCCAGGAAGCTGTGTGTAAATCAATCAATGAAGGCTGTGGCTAT 864  
661 QY GGTGTCAACTCCATCAAGTGTCTGCTGTTTCATGATGCTTATGGCAACCCAGGGAACA 720  
865 Db GGTGTCAACTCCATCAAGTGTCTGCTGTTTCATGATGCTTATGGCAACCCAGGGAACA 924  
721 QY CTCATCTTTGGAATCTCAGAAATGGCGCTATGCTACTGCTGGATGGAGACATGTTTGA 780  
925 Db CTCATCTTTGGAATCTCAGAAATGGCGCTATGCTACTGCTGGATGGAGACATGTTTGA 984  
781 QY CCTGTAAGTGAACATGTACAGACAGATCTGGCTCTCCACTGGACACTGGTCAAGTGAA 840  
985 Db CCTGTAAGTGAACATGTACAGACAGATCTGGCTCTCCACTGGACACTGGTCAAGTGAA 1044  
841 QY GTAAATGACAAACATTTCAAGTGTGAGCTGCCATTTGATAGACAGCTCCATCTCGG 900  
1045 Db GTAAATGACAAACATTTCAAGTGTGAGCTGCCATTTGATAGACAGCTCCATCTCGG 1104  
901 QY CCTCCTTACTTACCACTGCTGTTCCAGAGACCTTGCAGACCGACTCCTAAGAGTCCAT 960  
1105 Db CCTCCTTACTTACCACTGCTGTTCCAGAGACCTTGCAGACCGACTCCTAAGAGTCCAT 1164  
961 QY GGTGACCTGAGTGTGGTGTGCCAGTTTGTCAAATACCTTGATTCGTCACCAACCT 1020  
1165 Db GGTGACCTGAGTGTGGTGTGCCAGTTTGTCAAATACCTTGATTCGTCACCAACCT 1224  
1021 QY TGGCTGGAAGGAAATAGAGAGCCACCAAGAGCTTGGCTTCAACATCCAGTTAT 1080  
1225 Db TGGCTGGAAGGAAATAGAGAGCCACCAAGAGCTTGGCTTCAACATCCAGTTAT 1284  
1081 QY GGAGTCCATGTGACAGGACACAGCAAGTGGGAACAGAGAGCTTCCACCCCATCCAG 1140  
1285 Db GGAGTCCATGTGACAGGACACAGCAAGTGGGAACAGAGAGCTTCCACCCCATCCAG 1344  
1141 QY GAGTACATGGTACAGTGTGAAGAACATTTTCAGCTTCTCGCAGCAGAGATGCAAGTGGAT 1200  
1345 Db GAGTACATGGTACAGTGTGAAGAACATTTTCAGCTTCTCGCAGCAGAGATGCAAGTGGAT 1404  
1201 QY AAAAAAGAGTATCTGGCTACTGATGATCTTCTTGTAAAGGAGGCAAGCAAG 1260  
1405 Db AAAAAAGAGTATCTGGCTACTGATGATCTTCTTGTAAAGGAGGCAAGCAAG 1464  
1261 QY TACTCCAATTTATGATGATTAATCTATTTCTTGGTCAAGTGGAGTACCAAT 1320  
1465 Db TACTCCAATTTATGATGATTAATCTATTTCTTGGTCAAGTGGAGTACCAAT 1524  
1321 QY CGGTACACAGAAATTTCACTTCGGGGTGTGATCCTCGGTGATATACACTTTCTCACAGGCT 1380

1525 Db CGGTACACAGAAATTTCACTTCGGGGTGTGATCCTGGATATACACTTTCTCTCACAGGCT 1584  
1381 QY GACTTCTAGTGTCTACTTTTCAATCCAGCTGTGCGGGTGTCTTATGAATCATGCAA 1440  
1585 Db GACTTCTAGTGTGTACTTTTCAATCCAGCTGTGCGGGTGTCTTATGAATCATGCAA 1644  
1441 QY ACCCTGCATCCTGATCTCTCTGCGAACTTCATTTTGGATGACATCTACTATTTTGA 1500  
1645 Db ACCCTGCATCCTGATCTCTCTGCGAACTTCATTTTGGATGACATCTACTATTTTGA 1704  
1501 QY GGCCAAATGCCCAATCAGATTGCTTTTATCTCTCAAAACCTCGAACTGAAGAGAA 1560  
1705 Db GGCCAAATGCCCAATCAGATTGCTTTTATCTCTCAAAACCTCGAACTGAAGAGAA 1764  
1561 QY ATTCCAATGGAACCTCTGAGATATCATTTGGTGTGCTGGAACCATTTGGATGGTTATCT 1620  
1765 Db ATTCCAATGGAACCTCTGAGATATCATTTGGTGTGCTGGAACCATTTGGATGGTTATCT 1824  
1621 QY AAAGGTATCAACAGAAACTTTGGAACACAGGCTTATATCCCTCTCAAAAGTCCGAGAG 1680  
1825 Db AAAGGTATCAACAGAAACTTTGGAACACAGGCTTATATCCCTCTCAAAAGTCCGAGAG 1884  
1681 QY AAGATAGAAACAGTCAAGTATCCCATATATCCTGAAGCTGAAAAATAG 1728  
1885 Db AAGATAGAAACAGTCAAGTATCCCATATATCCTGAAGCTGAAAAATAG 1932

RESULT 4  
BD168516  
LOCUS 2008 bp DNA linear PAT 17-JAN-2003  
DEFINITION Cells producing antibody composition.  
ACCESSION BD168516  
VERSION BD168516.1 GI:27874328  
KEYWORDS WO 0231140-A/1.  
SOURCE Cricetulus griseus (Chinese hamster)  
ORGANISM Cricetulus griseus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
Cricetulus.  
REFERENCE 1 (bases 1 to 2008)  
AUTHORS Kanda,Y., Sato,M., Nakamura,K., Uchida,K., Shinkawa,T., Yamane,N.,  
Hosaka,E., Yamagaki,M. and Hanai,N.  
TITLE Cells producing antibody composition  
JOURNAL Patent: WO 0231140-A 1 18-APR-2002;  
KYOWA HAKKO KOGYO CO LTD  
COMMENT OS Cricetulus griseus (chinese hamster)  
PN WO 0231140-A/1  
PD 18-APR-2002  
PF 05-OCT-2001 WO 2001JP008804  
PR 06-OCT-2000 JP 00P 308526  
PI YUTAKA KANDA,MITSUO SATO,KAZUYASU NAKAMURA,KAZUHISA UCHIDA, PI  
TOYOHIDE SHINKAWA,NAOKO YAMANE,EMI HOSAKA,MOTOO YAMASAKI,NOBUO  
PI HANAI  
PC C12N5/10, C12F21/08, C07K16/00, A01K67/00, A61K39/395, C12N9/00, PC  
C12N15/52.  
PC G01N33/53  
CC Cells producing antibody composition  
FH Key Location/Qualifiers  
FT source 1..2008  
FT /organism='Cricetulus griseus (chinese hamster)',  
FT Location/Qualifiers  
FT 1..2008  
FT /organism='Cricetulus griseus'  
FT /mol\_type='genomic DNA'  
FT /db\_xref='taxon:10029'  
BASE COUNT 626 a 425 c 465 g 492 t  
ORIGIN

Query Match 91.7%; Score 1584; DB 6; Length 2008;  
Best Local Similarity 94.8%; Pred. No. 0;  
Matches 1638; Conservative 0; Mismatches 90; Indels 0; Gaps 0;



QTRNLGKHDEILRRRIENGAKELWFFLQSELKKLKNLEGNELQRLHDEFLDLGHHE  
RSIMTDLVLSQDAGWREKADLTTELVRITTYLONPKDCSKAKLVNINKGC  
GYCOLHHVVYCEMLAYCTORTLILLESONRYATGWETVPRPVSECTDRSGISTGH  
WSGEVKDNQVQVPLPVIDSLHPPPLPLAVPEDLADLVVHGDPAVWVVSQVXY  
LIRPQMLEKEIBEATKLGFGFVIGVHVRRTDKVGEAFHPIEBETMVHSEHFQL  
LARMQDKKRYLATDDPSLLKKAATKYPNYBFISNSISWAGLHNRYPENSLRGV  
ILDHFLSQADFLPCTFSQVRYAYEIMOTLHPDASANFSLDDIYFQGNANHQI  
AIYAHQPTADEIPMEPGDIIIGVAGNHWGYSKGVNRKLGRTGLYPSYKVRKLETVK  
YPTYPEABK"

BASE COUNT 533 a 367 c 418 g 441 t  
ORIGIN

Query Match 86.9%; Score 1501.4; DB 6; Length 1759;  
Best Local Similarity 91.8%; Pred. No. 0;  
Matches 1586; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 1 ATGGGGGATGACGTGGTCTCGGGTGGGATTATGCTCATCTCTTTTGGCTGGGGACC 60  
DB |||||  
QY 17 ATGGGGCCATGGACTGGTCTCGGGCTTGGATTATGCTCATCTCTTTTGGCTGGGGACC 76  
DB |||||  
QY 61 TTGTTATTTTATATAGTGGTCAATTTGGTTCGAGATAATGACCCCTGATCACTCCAGC 120  
DB |||||  
QY 77 TTGCTGTTTTATATAGTGGTCACTTTGGTACGAGATAATGACCCCTGATCACTCTAGC 136  
DB |||||  
QY 121 AGAGAACTCTCCAAGATTCTTGCAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 180  
DB |||||  
QY 137 CGAGAACTGTCCAAGATTCTGGCAAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 196  
DB |||||  
QY 181 AGGGGAATGGCTGAGTCTCTCGAATACAGAAAGCCCATATGACGGGGACAGCTACA 240  
DB |||||  
QY 197 AGGGGAATGGCGGAATCTCTCGGATACAGAAAGCCCATATTTGATCAGGGGCCAGCTATA 256  
DB |||||  
QY 241 GGAAGAGTCCGTGTTTGAAGACAGCTTGTGAAGCCAAAGACACAGATTGAAATTTAC 300  
DB |||||  
QY 257 GGAAGAGTACGCGTTTTGAAGAGCAGCTTGTGAAGCCAAAGACACAGATTGAAATTTAC 316  
DB |||||  
QY 301 AAGAAACAAGCTAGAAATGGTCTGGGGAAGGATCATGAAATCTTAAAGAGGAGGATTCAA 360  
DB |||||  
QY 317 AAGAAACAAGCAAGAAATGGTCTGGGGAAGGATCATGAAATCTTGAAGGAGGAGGATTGAA 376  
DB |||||  
QY 361 AATGAGCTAAAGAGCTCTGTTTTTCTCAAGCGAACTGAAGAAATTAAGCAATTTA 420  
DB |||||  
QY 377 AATGAGCTAAAGAGCTCTGTTTTTCTCAAGCGAACTGAAGAAATTAAGCAATTTA 436  
DB |||||  
QY 421 GAAGAAATGAACTCCAAAGACATGACAGTGAATTTCTTTGGATTAGGACACCATGAA 480  
DB |||||  
QY 437 GAAGAAATGAACTCCAAAGACATGACAGTGAATTTCTTTGGATTAGGACATCATGAA 496  
DB |||||  
QY 481 AGGTCTATCATGACAGATCTATCTACCTCAGTCAACAGATGGAGCAGGGGATTGGCGT 540  
DB |||||  
QY 497 AGGTCTAATGACGGATCTATCTACTCTCAGTCAAGATGGAGCAGGTGATTTGGCGG 556  
DB |||||  
QY 541 GAAAGAGGCGCAAGATCTGACAGAGTGGTCCAGCGGAGAAATTAATCTCCAGAAAT 600  
DB |||||  
QY 557 GAAAGAGGCGCAAGATCTGACAGAACTGGTTCAGCGGAGAAATTAATCTCCAGAAAT 616  
DB |||||  
QY 601 CCTAAGGCTGACAGAAAGCAGGAGCTGGTGTGATCATCAATTAAGGCTGGGCTAT 660  
DB |||||  
QY 617 CCCAAGGACTGACAGAAAGCCAAAGGCTGGTGTGATTAATCAACAAAGGCTGGGCTAT 676  
DB |||||  
QY 661 GGTGCTCAACTCCATCAGCTGGTCTACTGTTTCTATGTTTATGTCACCCAGCGAACA 720  
DB |||||  
QY 677 GGCTGTGAGTCCATCATGTGGTCTACTGTTCTATGTTGATATGGCACCCAGCGAACA 736  
DB |||||  
QY 721 CTCACTTTGGAATCTCAGAAATTTGGCGCTATGCTACTGTTGGATGGAGACTGTGTTTGA 780  
DB |||||  
QY 737 CTCACTTTGGAATCTCAGAAATTTGGCGCTATGCTACTGTTGGATGGAGACTGTATTTAGG 796  
DB |||||  
QY 781 CCTGTAAGTGAGACATGTACAGAGATCTGGCTCTCCATCTGACACTGGTCAAGTGA 840  
DB |||||  
QY 797 CCTGTAAGTGAGACATGTACAGAGATCTGGCTCTCCATCTGACACTGGTCAAGTGA 856  
DB |||||  
QY 841 GTAATGACAAACATTTCAAGTGGTGGCTGCCCTCCCATTTGTAGACAGCTCCATCCTCGG 900  
DB |||||

DB 857 GTGAAGCACAAAATGTTCAAGTGGTCCAGTTCCTCCATTTGTAGACAGTCTTTCATCCCCGT 916  
QY |||||  
DB 901 CCTCTTACTTACCACTGGCTGTTCCAGAGACCTTCAGACCGACTCCTTAAGAGTCCAT 960  
QY |||||  
DB 917 CCTCATATTTACCCCTTGGCTGTACCAAGAACCTTCGAGATCGACTTGTACGAGTGT 976  
QY |||||  
DB 961 GGTGACCCCTGCAAGTGTGGTGGTCTCCAGTTTGTCAAAATCTTGAATCGTCCCAACCT 1020  
QY |||||  
DB 977 GGTGACCCCTGCAAGTGTGGTGGTCTCAGTTTGTCAAAATCTTGAATCGGCCACAGCCT 1036  
QY |||||  
DB 1021 TGGCTGGAAGAAAGAAATAGAAAGACCAACAAAGCTTTGGCTTCAAAATCTCAGTTAT 1080  
QY |||||  
DB 1037 TGGCTAGAAAAGAAATAGAAAGACCAACAAAGCTTTGGCTTCAAAATCTCAGTTAT 1096  
QY |||||  
DB 1081 GGAGTCCATGTACAGACGACAGCAAAAGTGGAAACAGAGCCTTCCACCCCATCCAG 1140  
QY |||||  
DB 1097 GGAGTCCATGTACAGACGACAGCAAAAGTGGAAACAGAGCTGCTTCCATCCCATGAA 1156  
QY |||||  
DB 1141 GAGTACATGTGTACAGTTGAGAAACATTTTTCAGCTTCTCGCACGACAGAAATCAAGTGGAT 1200  
QY |||||  
DB 1157 GAGTACATGTGTGATGTTGAGAAACATTTTTCAGCTTCTCGCACGACAGAAATCAAGTGGAC 1216  
QY |||||  
DB 1201 AAAAAAGAGTATATCTGGCTACTGATGATCCTACTTTGTTTAAAGAGGCAAAAGCAAAAG 1260  
QY |||||  
DB 1217 AAAAAAGAGTATTTGGCCACAGATGACCCCTTCTTTTAAAGAGGCAAAAGCAAAAG 1276  
QY |||||  
DB 1261 TACTCCAAATATGAATTTATTTAGTGATACTCTATTTTCTTGGTCACTGAGTCAACAAT 1320  
QY |||||  
DB 1277 TACCCCAATATGAATTTATTTAGTGATACTCTATTTTCTTGGTCACTGAGTCAACAAT 1336  
QY |||||  
DB 1321 CGGTACACAGAAAATTCACATTCGGGGTGTGATCCTGGATATACACTTTTCTCTCACAGGCT 1380  
QY |||||  
DB 1337 CGATACACAGAAAATTCACATTCGGGGTGTGATCCTGGATATACACTTTTCTCTCACAGGCA 1396  
QY |||||  
DB 1381 GACTTCTAGTGTGTACTTTTTCATCCAGGCTGTGCGGGTGTGTTTGAATTAAGAAATCATGCAA 1440  
QY |||||  
DB 1397 GACTTCTAGTGTGTACTTTTTCATCCAGGCTGTGCGAGTGTGTTTGAATTAAGAAATCATGCAA 1456  
QY |||||  
DB 1441 ACCCTGATCCTGATGCTCTCGCAACTTCATTTCTTGGTCACTGATCACTATTTTGGGA 1500  
QY |||||  
DB 1457 ACATACATCCTGATGCTCTGCAAACTTCATTTCTTTAGATGATCACTATTTTGGG 1516  
QY |||||  
DB 1501 GGCCAAATGCCACAAATCAGATTGCTTTTATCTCACAACCTCGAACTGAAGAGGAA 1560  
QY |||||  
DB 1517 GGCCAGATGCCACAAATCAAAATTTGCTTATGCTTACCAACCCCGAACTGCAGATGAA 1576  
QY |||||  
DB 1561 ATTCCAATGGAACTGAGATATCATTTGTTGCTGGAAACCAATTTGGATGGTATTTCT 1620  
QY |||||  
DB 1577 ATTCCTATGGAACTGAGATATCATTTGTTGCTGGAAATCATTTGGATGGCTATTTCT 1636  
QY |||||  
DB 1621 AAAGGTATCAACAGAAATTTGGAACACAGGCTTATATCCCTCTCAAAAGTCCGAGAG 1680  
QY |||||  
DB 1637 AAAGGTGTCACAGAAATTTGGGAAGGACGGGCTTATATCCCTCTCAAAAGTTCGAGAG 1696  
QY |||||  
DB 1681 AAGATAGAAACAGTCAAGTATCCACATATCTCGAAGCTGAAAAATA 1727  
QY |||||  
DB 1697 AAGATAGAAACGTCAGTACCCCACTATCTCGAGCTGAGAAATA 1743  
QY |||||

## RESULT 6

BD087745  
LOCUS 1759 bp DNA linear PAT 27-AUG-2002  
DEFINITION Plant cells having function to add animal type sugar chain.  
ACCESSION BD087745  
VERSION BD087745.1 GI:22633355  
KEYWORDS JP 2001333787-A/1.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1759)  
AUTHORS Taniguchi, N., Seki, T. and Fujiyama, K.



DEFINITION Homo sapiens mRNA for N-Acetyl-beta-D-glucosaminide, complete cds.  
ACCESSION D89289  
VERSION 1  
KEYWORDS N-Acetyl-beta-D-glucosaminide; GDP-L-Fuc; alpha 1-6  
Fucosyltransferase; alphas-1-6 FucT.

SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (sites)  
AUTHORS Yanagidani, S., Uozumi, N., Ihara, Y., Miyoshi, E., Yamaguchi, N. and  
Taniguchi, N.

TITLE Purification and cDNA cloning of  
GDP-L-Fuc:N-acetyl-beta-D-glucosaminide:alpha-1-6 fucosyltransferase  
(alpha-1-6 FucT) from human gastric cancer MKN45 cells  
J. Biochem. 121 (3), 626-632 (1997)

JOURNAL 97279058  
MEDLINE 9133635  
PUBMED 2 (bases 1 to 2002)  
REFERENCE Yanagidani, S.

AUTHORS Purification and cDNA cloning of  
TITLE GDP-L-Fuc:N-acetyl-beta-D-glucosaminide:alpha 1-6  
Fucosyltransferase (alpha 1-6 FucT) from human stomach carcinoma  
MKN45 cells  
Unpublished

JOURNAL 3 (bases 1 to 2002)  
REFERENCE Taniguchi, N.  
AUTHORS Direct Submission  
TITLE Submitted (18-NOV-1996) Naoyuki Taniguchi, Osaka University Medical  
School, Department of Biochemistry, Yamadaoka 2-2, Suita, Osaka  
565, Japan (E-mail: proftani@biochem.med.osaka-u.ac.jp,  
Tel: 81-6-879-3420, Fax: 81-6-879-3429)

FEATURES  
source Location/Qualifiers  
1..2002  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
143..1870  
/note="GDP-L-Fuc; alpha 1-6 Fucosyltransferase; alpha-1-6  
FucT"  
/codon\_start=1  
/product="N-Acetyl-beta-D-glucosaminide"  
/protein\_id="BAA19764.1"  
/db\_xref="GI:2055307"

/translation="MRPTGSRWIMLILFANGTLLFYIGHLVRDNDHPDHSRELS  
KILAKLERKQONEDLRMAESLRIPEDGPAIGRVLEQLVRAKEQIENYKK  
QTRNGIKDHELRRIENGAKELFFLOSELKLNLEGNELQRAHDFLLDLGHE  
RSIMDLVYLSOTDAGDWREKEADLTQLVQRITVYLNQPKDCSKAKKLVNINKEG  
GYGCOLHVYCFMAYGTORTLILESONWRVATGCVETVPVSETCTDRSGISTGH  
WSGEVKNVQVVELPIVDSLHPRPYPLAVPEDLADRLVRVHGDPAVWVSOFVKY  
LIRPOPLEKEIEATKLGFKHPVIGVHVRRTDKVTEAFPHPIEEMVHVEHFQL  
LARRQVKKRYLATDPSLLKEAKTYPNYEFISDMSISWSAGLHNRVYENSURGV  
ILDHFLSQADPLVCTFSQVCRVAYEIMOTLHPDASANPHSLDDIYFYFGQNAHQI  
AIYAHQPRTADEIPMEPGDIIGVAGNHWGYSKGVNRKLGRTGLYPSYKVEKIEYVK  
YPTYPEAEK"

BASE COUNT 620 a 415 c 467 g 500 t  
ORIGIN  
Query Match 86.9%; Score 1501.4; DB 9; Length 2002;  
Best Local Similarity 91.8%; Pred. No. 0;  
Matches 1596; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 1 ATGCGGCGATGACTGGTTCCTGGCGTGGGATTATGCTCATTTCTTTTGGCTGGGGGACC 60  
Db 143 ATGCGGCGCATGACTGGTTCCTGGCGTGGGATTATGCTCATTTCTTTTGGCTGGGGGACC 202

QY 61 TTGTTATTTTATATAGTGGTTCATTTGGTTCAGATATAGACACCCCTGATCATCTCCAGC 120  
Db 203 TTGCTGTTTTATATAGTGGTTCATTTGGTTCAGATATAGACACCCCTGATCATCTAGC 262

QY 121 AGAGAACTCTCAAGATCTTTCGAAAGCTTGAACGCTTAAACAGCAAAATCAAGACTTG 180  
Db 263 CGAGAACTCTCAAGATCTTTCGAAAGCTTGAACGCTTAAACAGCAAAATCAAGACTTG 322

QY 181 AGCGAATGGCTGACTCTCTCCGAATACCAGAGGCCCCCATTTGACCGGGGACAGCTACA 240  
Db 323 AGCGAATGGCCGAATCTCTCCGATACCAGAGGCCCTATTGATCAGGGGCCAGCTATA 382  
QY 241 GGAAGAGTCCGTGTTTGAAGAAACAGCTTGTAAAGCCCAAGAACAGATGAAATTTAC 300  
Db 383 GGAAGAGTACCGCTTTTGAAGAGCAGCTTGTAAAGCCCAAGAACAGATGAAATTTAC 442  
QY 301 AAGAAAACAGCTAGAAATGGTCTCGGAAGGATCATGAAATCTTAAAGAGGAGATTGAA 360  
Db 443 AAGAAAACAGCCAGAAATGGTCTCGGAAGGATCATGAAATCTTAAAGAGGAGATTGAA 502  
QY 361 AATGAGAGTAAAGAGCTCTGGTTTTTCTACAAAGCCGAACCTGAAATTTAAAGCATTTA 420  
Db 503 AATGAGAGTAAAGAGCTCTGGTTTTTCTACAGAGTGAAATTTGAAAGAAATTTAAAGAACTTA 562  
QY 421 GAAGAAATGAATCTC3AAGACATGCAGATGAAATTTCTTTGGATTTTAGGACACCATGAA 480  
Db 563 GAAGAAATGAATCTC3AAGACATGCAGATGAAATTTCTTTGGATTTTAGGACATCATGAA 622  
QY 481 AGGTCTATCATGACAGATCTTACTCAGTCAAAACAGATGGAGAGGGGATTTGGCGT 540  
Db 623 AGGTCTATATGACG3ATCTATCTACTCAGTCAAGATGGAGAGGGGATTTGGCGG 682  
QY 541 GAAAAGAGGCCAAATCTCAGACAGATGGTCCAGCGGAGAAATAATATCTCCAGAAAT 600  
Db 683 GAAAAGAGGCCAAATCTCAGACAGATGGTCCAGCGGAGAAATAATATCTCCAGAAAT 742  
QY 601 CCTAAGAGCTGCAGCAAGCCAGAGAGCTGTGTGTAAATCAATTAAGAGCTGTGGCTAT 660  
Db 743 CCCAAGAGCTGCAGCAAGCCAAAGCTGTGTGTAAATCAATTAAGAGCTGTGGCTAT 802  
QY 661 GGTGTCAATCCATCAGCTGGTCTACTGTTTCTATGTTGCTTATGGCACCAGCGAACA 720  
Db 803 GGCTGTACGCTCCATATGTTGCTTCTATGTTGCTTATGTTGCTTATGGCACCAGCGAACA 862  
QY 721 CTCAATTTGGAATCTCAGAAATTTGGCGTATCTACTTGTGGATGGAGACTGTGTTAGA 780  
Db 863 CTCAATTTGGAATCTCAGAAATTTGGCGTATCTACTTGTGGATGGAGACTGTGTTAGA 922  
QY 781 CCTGTAGTGAAGATCTCAGACAGATCTGGCGCTCTCACTGTGGATGGAGACTGTGTTAGA 840  
Db 923 CCTGTAGTGAAGATCTCAGACAGATCTGGCGTCTCACTGTGGATGGAGACTGTGTTAGA 982  
QY 841 GTAATGAACAAACATTTCAAGTGGTCCAGCTCCCCATTTGTAGACAGCTCCATCCTCG 900  
Db 983 GTGAAGGACAAATATGTTCAAGTGGTCCAGCTCCCCATTTGTAGACAGCTCTCATCCCCGT 1042  
QY 901 CCTCTTACTTTACCACTGGCTGTTCCAGAGACCTTGCAGACCCGACTCCTTAAGAGTCCAT 960  
Db 1043 CCTCAATTTTCCCTTGGCTGTACAGAGACCTTGCAGAGTGCATTTGTACGAGTGCAT 1102  
QY 961 GGTGACCTCGAGCTGTGGTGGTGTCCAGTTTGTCAAATCTTGAATTCGTCCCAAACT 1020  
Db 1103 GGTGACCTCGAGCTGTGGTGGTGTCTCAGTTTGTCAAATCTTGAATTCGTCCCAAACT 1162  
QY 1021 TGGCTGGAAGAGGAAATGAAGAGACCCAGAGAGCTTTGGCTTCAACATCCAGTTAT 1080  
Db 1163 TGGCTGGAAGAGGAAATGAAGAGACCCAGAGAGCTTTGGCTTCAACATCCAGTTAT 1222  
QY 1081 GGAGTCCATGTCAGACGACAGCAAGAGTGGGAAACAGAGAGCTTCCACCCCATCCAG 1140  
Db 1223 GGAGTCCATGTCAGACGACAGCAAGAGTGGGAAACAGAGAGCTTCCATCCCATCCAG 1282  
QY 1141 GAGTACATGGTACACGTTGAAGAAACATTTTTCAGCTTCTCGACGAGAGATGCAAGTGGAT 1200  
Db 1283 GAGTACATGGTACGTTGAAGAAACATTTTTCAGCTTCTTCGACGAGAGATGCAAGTGGAC 1342  
QY 1201 AAAAAAGAGTATATCTGGCTATGATGATCTTCTTTTAAAGGAGGCAAGCAAG 1260  
Db 1343 AAAAAAGAGTATATCTGGCTATGATGATGATCTTCTTTTAAAGGAGGCAAGCAAG 1402







```
Db 1158 GGTGACCTGCAGTGTGGTGGTGTCTCAGTTTGTCAAATATCTTGATCCGCCACAGCCT 1217
Qy 1021 TGGCTGGAAGAGAAATAGAAAGCCACCAAGAGCTTGGCTTCAAACATCCAGTTATT 1080
Db 1218 TGGCTAGAAAAGAAATAGAGAGCCACCAAGAGCTTGGCTTCAAACATCCAGTTATT 1277
Qy 1081 GGAGTCCATGTGAGACGCACAGACAAAGTGGGAACAGAGCGCTTCCACCCCATCGAG 1140
Db 1278 GGAGTCCATGTGAGACGCACAGACAAAGTGGGAACAGAGCGCTTCCATCCCATGAA 1337
Qy 1141 GAGTACATGTTGACGTTGAAGACATTTTTCAGCTTCTCGACGACAGATCAAGTGGAT 1200
Db 1338 GAGTACATGTTGCAATTTTGAAGACATTTTTCAGCTTCTCGACGACAGATCAAGTGGAC 1397
Qy 1201 AAAAAACAGTATATCTGGCTACTGATGATCTTCTTCTTAAAGGAGGCAAGACAAAG 1260
Db 1398 AAAAAACAGTATTTTGGCCACAGATGACCTTCTTATTAAGGAGGCAAGACAAAG 1457
Qy 1261 TACTCCAAATATGAATTTATTAGTGAATCTTATTTCTTGGTCAAGTGGACTACACAAT 1320
Db 1458 TACCCCAATATGAATTTATTAGTGAATCTTATTTCTTGGTCAAGTGGACTGACCAAT 1517
Qy 1321 CGGTACACAGAAAATTCATCTCGGGGTGATCTGGATATACATTTCTCTCACAGGCT 1380
Db 1518 CGATACACAGAAAATTCATCTCGGGAGTGATCTGGATATACATTTCTCTCACAGCA 1577
Qy 1381 GACTTCTAGTGTGATCTTTTTCATCCAGGTCGTGGGTTGCTTATGAATCATGCAA 1440
Db 1578 GACTTCTAGTGTGATCTTTTTCATCCAGGTCGTGGGTTGCTTATGAATCATGCAA 1637
Qy 1441 ACCCTGCATCTCGATCGCTCTCGGAATCTTCCATCTTTTGGATGACATCTACTATTTTGA 1500
Db 1638 ACATCATCTCGATCGCTCTCGGAATCTTCCATCTTTTAGATGACATCTACTATTTTGG 1697
Qy 1501 GGCCAAATGCCCAATCAGATCTGCTGTTTATCTCTCAAAACCTCGAATCGAAGAGAA 1560
Db 1698 GGCCAGATGCCCAATCAATCAATTTTATGCTTCAACACCCGAACTGCAGATGAA 1757
Qy 1561 ATTCCAATGGAACTGGAGATATCATCTGTGGTGGTGAACCATTTGGATGGTTATTCT 1620
Db 1758 ATTCCAATGGAACTGGAGATATCATCTGTGGTGGTGAACCATTTGGATGGTTATTCT 1817
Qy 1621 AAAGGTATCAACAGAAAACTTGGAAAAACAGGCTTATATCCCTCTCAAAAGTCCGAGAG 1680
Db 1818 AAAGGTGTCAACAGGAATTTGGAGAGAGCGGCTATATCCCTCTCAAAAGTTCGAGAG 1877
Qy 1681 AAGATAGAAACAGTCAAGTATCCCAATATCTCTGAAGCTGAAAAATA 1727
Db 1878 AAGATAGAAACGTCAGTCAAGTATCCCAATATCTCTGAAGCTGAAAAATA 1924

RESULT 9
HSA539535
LOCUS HSA539535 mRNA 2796 bp linear PRI 03-FEB-2003
DEFINITION Homo sapiens mRNA for alpha6-fucosyltransferase (FUT8 gene), clone
CD-L1.
ACCESSION AJ539535
VERSION AJ539535.1 GI:28200807
KEYWORDS alpha6-fucosyltransferase; FUT8 gene.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Mollicone, R., Michalski, J.C., Bauvy, C., Calileau-Thomas, A.,
Candelier, J.J., Martinez-Duncker, I., Breton, C., Codogno, P. and
Oriol, R.
TITLE Splice variants of alpha6-fucosyltransferase are expressed early in
human embryogenesis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2796)
AUTHORS Oriol, R.
TITLE Direct Submission
```

```
JOURNAL Submitted (31-JAN-2003) Oriol R., U504, Inserm, 16 Av. Paul
Vaillant-Couturier, 94807, FRANCE
FEATURES
source
1..2796
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q23.2"
/clone="CD-L1"
/tissue_type="whole embryo"
/dev_stage="40-70 days"
/country="France:Paris"
1..2796
/gene="FUT8"
326..2053
/gene="FUT8"
/EC_number="2.4.1.68"
/function="adds fucose in alphanol,6 linkage to the internal
GlcNAc of the core chitobiose"
/evidence="experimental"
/product="alpha6-fucosyltransferase"
/protein_id="CAD62444.1"
/db_xref="GI:28200808"
/translation="MRPWTGSRWIMLIFAWGTLFLFYIGHVLVRDNDHPDHSRELS
KILAKRLKQNEDELRRMAESLRIPEGPIDOGPAIGRVRLBQVKAKEQIENYKK
OTRNLGKDEHILLRRRIENGAKELWFFLOSELKKNLEGNELORHAEFLDLGHHE
RSIMTDIYLSQTDGADWEKEAKDLTELVRITLYLQNPDKDSKAKLVNKNKGC
GYQLHVVYCFMIAQTORTLILESQNRWYATGWETFRPSETCTDKSGISTGH
WSCEPKDKVQVVELPIVDSLPRPPYPLVLAPEDLADRLVRVHGDPAVMVVSQFVK
LIRPQWLEKEIEBEATKLGKHPVIGVHRRTDKVGEAFAHPIDIEYMHVHESLFGV
LARMQVDRKRVLATDDPSLLKEAKTYPNYEPISONSISWSAGLHNRYPENSLRGV
ILIDHFLSQADFLVCTSPSCRVAVEMOTLHPDASANFSLDDIYVFGQNAHNOI
AIYAHQRIADEIPMEPGDIIIGVAGNHWGDSKGVNRKLGRTGLYPSYKVRKLETYK
YPTYPEAEK"
polya_signal
2766..2771
/gene="FUT8"
2777
polya_site
874 a 552 c 605 g 765 t
BASE COUNT
ORIGIN
Query Match 86.9%; Score 1501.4; DB 9; Length 2796;
Best Local Similarity 91.8%; Pred. No. 0;
Matches 1586; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
Qy 1 ATCGCGGCATGGACTGTTCTCTGGCGTTGGATTATGCTCATTTCTTTTGGCTGGGGACC 60
Db 326 ATCGCGGCATGGACTGTTCTCTGGCGTTGGATTATGCTCATTTCTTTTGGCTGGGGACC 385
Qy 61 TTGTTATTTATATAGTGGTCAATTTGGTTCGAGATATGACCACTGATCATCTCCAGC 120
Db 386 TTGCTGTTTTATATAGTGGTCACTTTGGTACGAGATATGACCACTGATCATCTTAGC 445
Qy 121 AGAGAACTCTCCAAGATTTCTTCAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 180
Db 446 CGAGAACTGTCCAAGATTTCTGCAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 505
Qy 181 AGCGCAATGGCTGAGTCTTCTCCGAATACCAAGAGCCCTTATGACGGGGACAGCTACA 240
Db 506 AGCGCAATGGCGGAATCTTCCCGATACCAAGAGCCCTTATGATCAGGGGCCAGCTATA 565
Qy 241 GGAAGAGTCCGTGTTTGTAGAAAGACAGCTTGTTAAGGCCAAAGAACAGATGAAAATTAC 300
Db 566 GGAAGAGTACGCGTTTGTAGAAAGACAGCTTGTTAAGGCCAAAGAACAGATGAAAATTAC 625
Qy 301 AAGAAACAGCTAGAAATGGTCTCGGGAAGGATCATGAAATCTTAAAGAGGAGATTGAA 360
Db 626 AAGAAACAGCTAGAAATGGTCTCGGGAAGGATCATGAAATCTTAAAGAGGAGATTGAA 685
Qy 361 AATGGAGCTAAGAGCTCTGTTTCTTCTCAAAAGCGAATCTGAAGAAATTAAGCATTTA 420
```

Db 686 AATGAGCTAAGAGCTCTGGTTTTCTCTCAGAGTGAATTGAAGAAATTAAGAACTTA 745  
Qy 421 GAAGGAATGAATCTCAAGACATGAGATGAAATCTTTTGGATTTAGGACACCATGAA 480  
Db 746 GAAGGAATGAATCTCAAGACATGAGATGAAATCTTTTGGATTTAGGACATCATGA 805  
Qy 481 AGGTCATCATGACAGATCTATACCTCAGTCAACAGATGGAGAGGGGATTTGGCT 540  
Db 806 AGGTCATCATGACGGATCTATACCTCAGTCAACAGATGGAGAGGGGATTTGGCGG 865  
Qy 541 GAAGAAGAGCCCAAGATCTGACAGAGCTGGTCCAGCGGAGAAATAACATATCTCCAGAT 600  
Db 866 GAAGAAGAGCCCAAGATCTGACAGAGCTGGTCCAGCGGAGAAATAACATATCTCCAGAT 925  
Qy 601 CTTAAGGAGCTGACGAAAGCCAGGAGCTGGTGTAAACATCAATAAAGGCTGTGGCTAT 660  
Db 926 CCCAAGGAGCTGACGAAAGCCCAAGCTGGTGTATATCAACAAAGGCTGTGGCTAT 985  
Qy 661 GGTTCCTCAATCCTACATGCTGGTCTATCTGTTTCATGATTTATGGCAACCCAGCAACA 720  
Db 986 GGCTGTGAGCTCCATCATGCTGCTTACTGCTTCATGATTTGCATATGGCACCCAGCAACA 1045  
Qy 721 CTCATCTTGGATCTCAGAAATGGCGCTATGCTACTGGTGGATGGAGACTGTATTTAGG 780  
Db 1046 CTCATCTTGGATCTCAGAAATGGCGCTATGCTACTGGTGGATGGAGACTGTATTTAGG 1105  
Qy 781 CTTGTAAGTGAGACATGTACAGACAGATCTGGCCCTCTCCACTGGACACTGGTCAAGTCAA 840  
Db 1106 CTTGTAAGTGAGACATGTACAGACAGATCTGGCATCTCCATCGACACTGGTCAAGTCAA 1165  
Qy 841 GTAATGACAAAGAAATGAAGAGCCACCAAGAGCTTTGGCTTCAACATCCAGTTAT 900  
Db 1166 GTGAAGGACAAAGAAATGAAGAGCCACCAAGAGCTTTGGCTTCAACATCCAGTTAT 1225  
Qy 901 CCTCTTACTTACACTGGCTGTTCCAGAGACCTTGCAGACCGACTCCTAAGAGTCCAT 960  
Db 1226 CCTCCATATTTACCTTGGCTGTACCAAGAGACCTTGCAGATCGACTTGTACGAGTGCAT 1285  
Qy 961 GGTGACCTTGAGTGTGGTGGTGTCCAGAGTTTGTCAATAATCTTCAATTCGTCACCAAGCT 1020  
Db 1286 GGTGACCTTGAGTGTGGTGGTGTCTCAGTTTGTCAATAATCTTCAATTCGTCACCAAGCT 1345  
Qy 1021 TGGCTGGAAGAAATGAAGAGCCACCAAGAGCTTTGGCTTCAACATCCAGTTAT 1080  
Db 1346 TGGCTGGAAGAAATGAAGAGCCACCAAGAGCTTTGGCTTCAACATCCAGTTAT 1405  
Qy 1081 GGAGTCCATGTGACAGCGACACAGAAAGTGGGAAGCAGAGCGCTTCCACCCCATCGAG 1140  
Db 1406 GGAGTCCATGTGACAGCGACACAGAAAGTGGGAAGCAGAGCTTCCATCCCATTTGAA 1465  
Qy 1141 GAGTACATGTTACAGCTTTGAAGAACATTTTCAAGCTTCTCGACGAGAGTCAAGTGGAT 1200  
Db 1466 GAGTACATGTTGATGTTGAAGAACATTTTCAAGCTTCTCGACGAGAGTCAAGTGGAT 1525  
Qy 1201 AAAAAAGAGTATATCTGGCTACTGATGATCTTCTTTGTTAAAGGAGGCAAGCAAG 1260  
Db 1526 AAAAAAGAGTATATTTGGCCACAGATGACCCCTCTTTTATTAAAGGAGGCAAGCAAG 1585  
Qy 1261 TACTTCAATATGAATTTATAGTGAATCACTTATTTCTTGGTCACTGGACTACACAT 1320  
Db 1586 TACCCTCAATATGAATTTATAGTGAATCACTTATTTCTTGGTCACTGGACTACACAT 1645  
Qy 1321 CGGTACACAGAAATTCACCTCGGGGTGTGATCCTCGGATATACATTTCTCTCACAGCT 1380  
Db 1646 CGATACACAGAAATTCACCTCGGGGTGTGATCCTCGGATATACATTTCTCTCACAGCT 1705  
Qy 1381 GACTTTCTAGTGTGATCTTTTTCATCTCCAGGCTGTGCGGGTGTGATGAAATCATGAA 1440  
Db 1706 GACTTTCTAGTGTGATCTTTTTCATCTCCAGGCTGTGCGAGTGTGATGAAATCATGAA 1765  
Qy 1441 ACCCTGATCTGATGCTCTCGGAGCTTCCATTTCTTGGATGATCACTACTATTTTGGG 1500  
Db 1766 ACACCTACATCTGATGCTCTCGGAGCTTCCATTTCTTGGATGATCACTACTATTTTGGG 1825

Qy 1501 GGCCAAATGCCACAATCAGATTCGTTTATCTCACAACCTCGACTGAAGAGAA 1560  
Db 1826 GGCCAGAATGCCACAATCAGATTCGTTTATCTCACAACCTCGACTGAAGAGAA 1885  
Qy 1561 ATTCCAATGGAACCTGGAGATATCATTTGGTGTGCTGGAACCATTTGGATGGTATTCT 1620  
Db 1886 ATTCCAATGGAACCTGGAGATATCATTTGGTGTGCTGGAACCATTTGGATGGTATTCT 1945  
Qy 1621 AAAGGTATCAACAGAAACTTTGGAAAAACAGGCTTATATCCCTCTCAAAAAGTCCGAGAG 1680  
Db 1946 AAAGGTATCAACAGAAACTTTGGAAAAACAGGCTTATATCCCTCTCAAAAAGTCCGAGAG 2005  
Qy 1681 AAGTAGAAGACAGTCAAGTATCCACATATCTCTGAAGCTGAAAAATA 1727  
Db 2006 AAGTAGAAGACAGTCAAGTATCCACATATCTCTGAAGCTGAAAAATA 2052

RESULT 10  
LOCUS HSY17976 2898 bp mRNA linear PRI 17-SEP-1998  
DEFINITION Homo sapiens mRNA for glycoprotein 6-alpha-L-fucosyltransferase, transcript B1.  
ACCESSION Y17976  
VERSION Y17976.1 GI:3451262  
KEYWORDS FUT8 gene; glycoprotein 6-alpha-L-fucosyltransferase.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE 1  
AUTHORS Calileau,A., Balanzino,L., Candellier,J.J., Oriol,R. and Mollicone,R.  
TITLE Differential splice variants of human FUT8 embryonic cDNA  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2898)  
AUTHORS Calileau,A.L.V.  
TITLE Direct Submission  
JOURNAL Submitted (19-AUG-1998) A.L.V. Calileau, INSERM U178, 16 av P.V. Couturier, 94807 Villejuif Cedex, FRANCE  
COMMENT Related entry: D89289.  
FEATURES  
source  
1..2898  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/map="14q23"  
/clone="FUT8 transcript B1"  
/tissue\_type="whole embryo"  
/dev\_stage="40-70 days"  
1..2898  
/gene="FUT8"  
1..101  
/gene="FUT8"  
/note="B"  
102..200  
/gene="FUT8"  
/note="C"  
201..2156  
/gene="FUT8"  
/note="D"  
429..2156  
/gene="FUT8"  
/EC\_number="2.4.1.68"  
/note="allele 1"  
/codon\_start=1  
/product="glycoprotein 6-alpha-L-fucosyltransferase"  
/protein\_id="CAA76985.1"  
/db\_xref="GI:3451263"  
/db\_xref="SPTREMBL:O00235"  
/translation="MRPTGSRWIMLILFAWGTLLFVIGHLVRDNDHPDHSRELS  
KILALRLKQNDLRLMAESLRIPGPDQGPALGRVRLVEQLVKAKQIENYKK

QTRNLGKHDEILRRRIENGAKELWFFLQSELKKLKNLEGNLORHADEFLLDLGHE  
RSIMDLVYLSDGAGDWREKADLTTELVRRIITVLPKDCSKAKLVNINKGC  
GYCOLHVVYCFMAYACTORTLILESONWRYATGWETVFRPVSETCTDRSGISTGC  
WSGVBKNQVOVELPIVDLSLHPRPPIPLAVPEDLADRLVRVHGDPAVWVSOFVKY  
LIRPQMLEKIEIEATKLGPKFPIGVHVRRTDKVGEAAPFIEETVMVHEHPQL  
LARMQVDRKRVYLATDDPSLLKBAKTKPNYEFISONSISWSAGLHNRVTENSLRGV  
ILDIFLSQADFLVCTTSSQVQVAYEIMOTLPDASANKFSLDDIYFQGNANHOI  
AIYAHQPRDDEIPMEPGDIIGVAGNHDWYSGKGNRKLGRGLYPSVKVREKJETVK  
YPTYPEAKK"

## variation

2218  
/gene="FUT8"  
/note="II"  
/frequency=".5"  
/replace="c"  
2758..2763  
/gene="FUT8"  
/note="putative"  
2770..2775  
/gene="FUT8"  
/note="putative"  
2869..2874  
/gene="FUT8"  
/note="putative"  
BASE COUNT 881 a 589 c 641 g 787 t  
ORIGIN

Query Match 86.9%; Score 1501.4; DB 9; Length 2898;

Best Local Similarity 91.8%; Pred. No. 0;  
Matches 1586; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 1 ATGCGGGCATGGACTGGTCTCTGGCGTTGGGATTATGCTCAITCTTTTTTGGCTGGGGGACC 60  
DB 429 ATGCGGCATGGACTGGTCTCTGGCGTTGGGATTATGCTCAITCTTTTTTGGCTGGGGGACC 488  
QY 61 TTGTTATTTTATATAGTGGTCAITTTGGTTCGAGATAATGACCCCTGATCACTCCAGC 120  
DB 489 TTGCTGTTTATATAGTGGTCAITTTGGTTCGAGATAATGACCCCTGATCACTCCAGC 548  
QY 121 AGAGAACTCTCAAGATTTCTGCAAGCTTGAACGCTTAAACACGAAATGAAGACTTG 180  
DB 549 CGAGAACTGTCCAGATTTCTGCAAGCTTGAACGCTTAAACACGAAATGAAGACTTG 608  
QY 181 AGGCGAATGGTGTAGTCTCTCGAATACAGAAAGGCCCATTTGACCGGGGACAGCTACA 240  
DB 609 AGGCGAATGGCGAATCTCTCGGATACAGAAAGGCCCATTTGATCAGGGGCCAGCTATA 668  
QY 241 GGAAGAGTCCGTGTTTATAGAAAGACAGCTTTTAAAGCCAAAGAACACAGATGAAATPAC 300  
DB 669 GGAAGAGTACGCGTTTATAGAAAGACAGCTTTTAAAGCCAAAGAACACAGATGAAATPAC 728  
QY 301 AAGAAACAAGCTAGAAATGGTCTGGGGAAGGATCATGAATCTTAAAGAGGAGGATTCGA 360  
DB 729 AAGAAACAAGCTAGAAATGGTCTGGGGAAGGATCATGAATCTTAAAGAGGAGGATTCGA 788  
QY 361 AATGAGCTTAAAGAGCTCTGGTGTCTTCTCAAGCGAACTGAAGAAATTAAGCAATTA 420  
DB 789 AATGAGCTTAAAGAGCTCTGGTGTCTTCTCAAGCGAACTGAAGAAATTAAGCAATTA 848  
QY 421 GAAGAAATGAATCTCAAGAGCATCGAGATGAAATCTTTTGGATTTAGGACACCATGAA 480  
DB 849 GAAGAAATGAATCTCAAGAGCATCGAGATGAAATCTTTTGGATTTAGGACACCATGAA 908  
QY 481 AGGTCTATCATGACAGATCTATACCTCAGTCAAAAGATGAGAGGAGGATTTGGCGT 540  
DB 909 AGGTCTATATGACGGATCTATACCTCAGTCAAAAGATGAGAGGAGGATTTGGCGG 968  
QY 541 GAAAGAGGCGCAAGATCTGACAGCTGGTCCAGCGGAGAAATTAACATATCTCCAGAAAT 600  
DB 969 GAAAGAGGCGCAAGATCTGACAGCTGGTCCAGCGGAGAAATTAACATATCTCCAGAAAT 1028  
QY 601 CCTAAGGACTGCAGCAAGCCAGGAGCTGGTGTGTAACATCAATPAAAGGCTGGCTTAT 660  
DB 1029 CCCAAGGACTGCAGCAAGCCAGGAGCTGGTGTGTAACATCAATPAAAGGCTGGCTTAT 1088

QY 661 GGTGTCAACTCCAAACACGTGGTCTACTGTCTTTCATGATTGCTTATGGCACCACCGGAACA 720  
DB 1089 GGTGTCACTCCAAACATGTGGTCTACTGTCTTTCATGATTGCTTATGGCACCACCGGAACA 1148  
QY 721 CTCACTTTGGAATCTCAGAAATTTGGCGCTATGCTACTCTGGTGGATGGGAGACTGTGTTAGA 780  
DB 1149 CTCACTTTGGAATCTCAGAAATTTGGCGCTATGCTACTCTGGTGGATGGGAGACTGTGTTAGA 1208  
QY 781 CCTGTAAGTGAGACATCTACAGACAGATCTGGCCTCTCCACTGGACACTGTGTCAGGTGAA 840  
DB 1209 CCTGTAAGTGAGACATCTACAGACAGATCTGGCCTCTCCACTGGACACTGTGTCAGGTGAA 1268  
QY 841 GTAAATGACAAAAACATTTCAAGTGGTGGCTCCCAATTTGTAGACAGCTCCATCTCTCGG 900  
DB 1269 GTGAAGGACAAAATGTTCAAGTGGTGGCTCCCAATTTGTAGACAGCTCTTCACTCCCGT 1328  
QY 901 CCTCTTACTTACCACTGGCTGTTCCAGAGAGACTTTGGAGACCGACTCTTAAGAGTCCAT 960  
DB 1329 CCTCATATTTACCCCTTGGCTGTACCAAGAGACCTCGCAGATCGACTTGTACGAGTGCAT 1388  
QY 961 GGTGACCTCGAGTGTGGTGGTGTCCAGCTTTGCTCAAACTACTTGAATTCGTCACAACCT 1020  
DB 1389 GGTGACCTCGAGTGTGGTGGTGTCTCAGTGTGTCAAACTACTTGAATTCGTCACAACCT 1448  
QY 1021 TGGCTGGAAGAGAAATAGAAAGCCCAAGAAAGCTTTGGCTTCAAACTACTTGAATTCGTCACAACCT 1080  
DB 1449 TGGCTGGAAGAGAAATAGAAAGCCCAAGAAAGCTTTGGCTTCAAACTACTTGAATTCGTCACAACCT 1508  
QY 1081 GGAGTCCATGTACAGACGACAGAAAGTGGGAAACAGAGAGCTTCCACCCCATCAG 1140  
DB 1509 GGAGTCCATGTACAGACGACAGAAAGTGGGAAACAGAGAGCTTCCACCCCATCAG 1568  
QY 1141 GAGTACATGTGTACACCTTTGAAGAACTTTTCAAGCTTCTCGCAGCAGAAATGCAAGTGCAT 1200  
DB 1569 GAGTACATGTGTGTGCAATTTGAAGAACTTTTCAAGCTTCTCGCAGCAGAAATGCAAGTGCAT 1628  
QY 1201 AAAAAAGAGTATATCTGGCTACTGATGATCCTACTTTGTTTAAAGAGGAGGCAAGCAAG 1260  
DB 1629 AAAAAAGAGTATATTTGGCCACAGATGACCCCTTTTATTAAGAGGAGGCAAGCAAG 1688  
QY 1261 TACTCCAATATGAATTTATTTAGTATAACTCTATTTCTTGGTCACTTCTCTCACAAAT 1320  
DB 1689 TACCCCAATATGAATTTATTTAGTATAACTCTATTTCTTGGTCACTTCTCTCACAAAT 1748  
QY 1321 CGGTACACAGAAATTTCACTTCCGGGTGTGATCTGAGATATACATTTCTCTCACAGCT 1380  
DB 1749 CGGTACACAGAAATTTCACTTCCGGGTGTGATCTGAGATATACATTTCTCTCACAGCT 1808  
QY 1381 GACTTTCTAGTGTCTACTTTTTCATCCAGGTCTGTGGGTTGCTTATGAATCATGCA 1440  
DB 1809 GACTTTCTAGTGTGTACTTTTTCATCCAGGTCTGTGGGTTGCTTATGAATCATGCA 1868  
QY 1441 ACCCTGCTATCTGATGCTCTCGAACTTCCATTTCTTTGGATGATCATCTATTTTGG 1500  
DB 1869 ACCTTACATCTGATGCTCTCGAACTTCCATTTCTTTAGATGATCATCTATTTTGG 1928  
QY 1501 GGCCAAAATGCCCAATCATGATTTGCTTTTATCTCACAACCTCGAACTGAGAGGAA 1560  
DB 1929 GGCCAAAATGCCCAATCAAATTTGCCATTTATGCTCACAACCTCGAACTGAGAGGAA 1988  
QY 1561 ATTCCAATGGAACCTGGAGATATCATTTGGTGTGGTGGGAAACCATTTGGATGTTTATCT 1620  
DB 1989 ATTCCAATGGAACCTGGAGATATCATTTGGTGTGGTGGGAAATCATTTGGATGTTTATCT 2048  
QY 1621 AAAGGTATCAACAGAAACTTTGGAAAAACAGGCTTATATCTCTCTCAAAAGTCCGAGAG 1680  
DB 2049 AAAGGTATCAACAGAAACTTTGGGAAAGGAGCGGCTTATATCTCTCTCAAAAGTCCGAGAG 2108  
QY 1681 AAGATAGAAACAGTCAAGTATCCACATATCTCTGAAAGCTGAAAAATA 1727  
DB 2109 AAGATAGAAACAGTCAAGTATCCACATATCTCTGAGGCTGAGAAATA 2155

RESULT 11	HSAS539536	2902 bp	mRNA	linear	PRI 03-FEB-2003
LOCUS	HSAS539536				
DEFINITION	Homo sapiens mRNA for alpha6-fucosyltransferase (FUT8 gene), clone DCD-12.				
ACCESSION	AJ539536				
VERSION	AJ539536.1	GI:28200809			
KEYWORDS	alpha6-fucosyltransferase; FUT8 gene.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1	Mollicone, R., Michalski, J.C., Bauvy, C., Cailletau-Thomas, A., Candelier, J.J., Martinez-Duncker, I., Breton, C., Codogno, P. and Oriol, R.			
AUTHORS		Splice variants of alpha6-fucosyltransferase are expressed early in human embryogenesis			
TITLE		Unpublished			
JOURNAL		2 (bases 1 to 2902)			
REFERENCE		Direct Submission			
AUTHORS		Submitted (31-JAN-2003) Oriol R., U504, Inserm, 16 Av. Paul Vaillant-Couturier, 94807, FRANCE			
JOURNAL		Location/Qualifiers			
FEATURES		1..2902			
source		/organism="Homo sapiens"			
		/mol_type="mRNA"			
		/db_xref="taxon:9606"			
		/chromosome="14"			
		/map="14q23.2"			
		/clone="CD-12"			
		/tissue type="whole embryo"			
		/dev stage="40-70 days"			
		/country="France:Paris"			
gene		1..2902			
		/genes="FUT8"			
CDS		326..2053			
		/genes="FUT8"			
		/EC number="2.4.1.68"			
		/function="adds fucose in alpha1,6 linkage to the internal GlcNAc of the core chitobiose"			
		/codon_start=1			
		/evidence=experimental			
		/product="alpha6-fucosyltransferase"			
		/protein_id="CAD62445.1"			
		/db_xref="GI:28200810"			
		/translation="MRPWTGSRWIMLILFAWGTLLFVIGGHLVRDNDHPHSSRELS KILAKLERLKQNEDLRMASLSRIPEGPIDQGAIGRVRVLEQLVKAKEQIENYKK QRTNGKGDHLEILRRRIENGAKELMFFLOSELKKLKNLEGNELQRHDEFLLDLGHHE RSIMTDLYLTDGADGWREKAKDLTELVRITLYLQNPDKCSKAKLVCVNIKNGC GYCVQLHVYCFMIAVTGTRTLILLESQNRWYATGQWETVPRPVSETCTDRSGISTGH WSGEVKDNVQVPELPIVDLSLHPRPPYPLAVPEDLADRLVRHGDPAVWVWSQFVKY LIRPWPBKEKEEATKGLGPKHPVIGVHVVRTDKVGTAAAFHPIDBYVMVHREHFOL LARMPQKRVYLATDPSLLKEAKTKYPNYEPISDNSISWAGLHNRVTENSLRGV ILDIHPSQADFVLCVSQVCRVAIEIMQILHPDASFNHSLDDIIFYFGQNAHQI AYIAQHPRTADSIPEMPGDIIGVAGNHWGDSKGVNKRKLGRITGLPSYKVRKIEITYK YTPPEAEK"			
		2853..2858			
		/genes="FUT8"			
		2887			
		/genes="FUT8"			
BASE COUNT	907 a	561 c	627 g	807 t	
ORIGIN					
Query Match		86.9%;	Score 1501.4;	DB 9;	Length 2902;
Best Local Similarity		91.8%;	Pred. No. 0;		
Matches 1586;		Conservative	0;	Mismatches	141;
				Indels	0;
				Gaps	0;
Qy	1	ATCGCGGCGATGGACTGGTTCCTGGCGTTGGATTATGCTCATTTCTTTTTCCTGGGGAC	60		
	326	ATCGCGCCATGACTGGTTCCTGGCGTTGGATTATGCTCATTTCTTTTTCCTGGGGAC	385		

QY 1141 GAGTACATGGTACAGCTTTGAAGAACATTTTTCAGCTTCTCGCAGCAGAGTCAAGTGGAT 1200  
Db 1466 GAGTACATGGTGCATGTTGAAGAACATTTTTCAGCTTCTTCGACGAGATGCAAGTGAC 1525  
QY 1201 AAAAAAGAGTATATCTGGCTACTGATGATCTCTTTGTTAAAGGAGGCAAGACAAAG 1260  
Db 1526 AAAAAAGAGTATATTTGGCCACAGATGACCCCTCTTTTATTAAGGAGGCAAGACAAAG 1585  
QY 1261 TACTCCAAATGATATTTATTAGTATGATTAATCTCTTTCTGTCAGCTGGACTACACAAT 1320  
Db 1586 TACCCCAATGATATTTATTAGTATGATTAATCTCTTTCTGTCAGCTGGACTGACAAAT 1645  
QY 1321 CGGTACACAGAAAATTCATCTCGGGGTGTGATCCCTGGATATACATTTCTCTCACAGGCT 1380  
Db 1646 CGATACACAGAAAATTCATCTCGTGGAGTATCCCTGGATATACATTTCTCTCTCAGGCA 1705  
QY 1381 GACTTTCTAGTGTGATCTTTTTCATCCAGGTCCTGCGGTTGCTTATGAATCATGCAA 1440  
Db 1706 GACTTCTCTAGTGTGATCTTTTTCATCCAGGTCCTGCGGTTGCTTATGAATCATGCAA 1765  
QY 1441 ACCCTGCACTCGTCTGCGGACCTTCCATCTTTTGGATGACATCTACTATTTTGGGA 1500  
Db 1766 ACATACATCTGATGCGCTCTGCAAACTTCCATCTTTTAGATGACATCTACTATTTTGGG 1825  
QY 1501 GGCCAAAATGCCCAATCAGATTTGCTGTTTATCTCTCAAACTCGAACTGAAGAGAA 1560  
Db 1826 GGCCAGATGCCCAATCAATTCGCAATTTATGCTTCAACACCCGCACTGCAGATGAA 1885  
QY 1561 ATTCCAAATGGAACTCGGAGATATCATTTGGTGTGGCTGGAACCAATGGGATGTTTCT 1620  
Db 1886 ATTCCATGGAACTCGGAGATATCATTTGGTGTGGCTGGAACCAATGGGATGTTTCT 1945  
QY 1621 AAGGTATCAACAGAAAATTTGGAAAACAGGCTTATATCTCTCAAACTGAAGTCCGAG 1680  
Db 1946 AAGGTGTCAACAGAAAATTTGGAAAACAGGCTTATATCTCTCAAACTGAAGTCCGAG 2005  
QY 1681 AAGTAGAAAACAGTCAAGTATCCCATATCTCTGAAGCTGAAAATA 1727  
Db 2006 AAGTAGAAAACAGTCAAGTATCCCATATCTCTGAAGCTGAAAATA 2052

RESULT 12  
HSY17977  
LOCUS HSY17977 2992 bp mRNA linear PRI 17-SEP-1998  
DEFINITION Homo sapiens mRNA for glycoprotein 6-alpha-L-fucosyltransferase, transcript B2.  
ACCESSION Y17977  
VERSION Y17977.1 GI:3451264  
KEYWORDS FUT8 gene; glycoprotein 6-alpha-L-fucosyltransferase.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Cailleau, A., Balanzino, L., Candelier, J.J., Oriol, R. and Mollicone, R.  
TITLE Differential splice variants of human FUT8 embryonic cDNA  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2992)  
AUTHORS Cailleau, A.L.V.  
TITLE Direct Submission  
JOURNAL Submitted (19-AUG-1998) A.L.V. Cailleau, INSERM U178, 16 av P.V. Coururier, 94807 Villejuif Cedex, FRANCE  
COMMENT Related entry: D89289.  
FEATURES  
source  
1..2992  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/isolate="transcript B2"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/map="14q23"  
/tissue\_type="whole embryo"

gene  
exon  
exon  
exon  
CDS  
variation  
polyA\_signal  
polyA\_signal  
polyA\_signal  
BASE COUNT  
ORIGIN  
Query Match  
Best Local Similarity  
Matches 1586; Conservative 0; Mismatches 141; Indels 0; Gaps 0;  
QY 1 ATGGCGGCATGGACTGTTCTCTGGCGTTGGATTATGCTCATTTCTTTTGGCTGGGGACC 60  
Db 429 ATGGCGGCATGGACTGTTCTCTGGCGTTGGATTATGCTCATTTCTTTTGGCTGGGGACC 488  
QY 61 TTGTTATTTTATAGTGGTGCATTTGTTTCGAGATATGACCCTGATCATCTCCAGC 120  
Db 489 TTGCTGTTTATAGTGGTGCATTTGTTTCGAGATATGACCCTGATCATCTCTAGC 548  
QY 121 AGAGACTCTCAAGATTTCTTGAAGCTTGAACGCTTAAACAGCAAAATCAAGACTTG 180  
Db 549 CGAGAACTGTCCAAGATTTCTGCAAGCTTGAACGCTTAAACAGCAAAATCAAGACTTG 608  
QY 181 AGGCAATGGCTGAGTCTCTCCGAATACCAAGGCCCTTATGACAGGGGACAGCTACA 240  
Db 609 AGGCAATGGCCGAATCTCTCCGAATACCAAGGCCCTTATGATCAGGGCCAGCTATA 668  
QY 241 GGAAGAGTCCGTGTTTATGAAGAACAGCTTGTATAGGCCCAAGAACAGATTTGAAATTTAC 300  
Db 669 GGAAGAGTACGCTTTTATGAAGAGCAGCTTGTATAGGCCCAAGAACAGATTTGAAATTTAC 728  
/dev stage="40-70 days"  
1..2992;  
/gene="FUT8"  
1..101  
/gene="FUT8"  
/note="H"  
102..200  
/gene="FUT8"  
/note="C"  
201..2156  
/gene="FUT8"  
/note="P"  
429..2156  
/gene="FUT8"  
/EC number="2.4.1.68"  
/note="allele 1"  
/codon\_start=1  
/product="glycoprotein 6-alpha-L-fucosyltransferase"  
/protein\_id="CAA76986.1"  
/db\_xref="GI:3451265"  
/db\_xref="SPTREMBL:O00235"  
/translation="MRPTSGSRWIMLILFAWGTLLEFVIGHLVRDNDHPDHSRELS  
KILAKRLQONEDLRMAESLRIPGPIIDQGPIDQGPAGRVLRVLEQLVAKQIENYKK  
QTRNLGKDHEILRRRIENGAKELWFFLQSELKKNLEGNLQKHADEFLLDLGHHH  
RSIMTDLYLSQTDAGDREKEADTELVRITLYLQNPDCSKAKLVNINKGC  
GYGOLHHVYCFMIAVGTQRTILLESQNRVATGWETVRPVSETCTDRSGTSGH  
WSGFKDKQVQVVELPIVDSLHPRPPLPLAVPEDLADRLVRVHGDPAVMVVEHFQ  
LIRQPLKEIEBEATKLGFKHPVIGVHVRRTDKVGEAFHPIEEMVHVEHFQ  
LARMQVDDKRVLATDDPSLLKEAKTKPNYEFSNWSAGLNHNRITENSURGV  
LDLHFLSQADFLVCTSSQVCRVAYEIMQTLHPDASANFSLDDIYFPGQNAHQI  
AIVAHQPTADEIPMEPDIIGVAGNHWGYSKGNRKLGRGLTGLYPSYKVRKREITVK  
VPTYPEAEK"  
2218  
/gene="FUT8"  
/note="11"  
/frequency="0.5"  
/replace="c"  
2758..2763  
/gene="FUT8"  
/note="putative"  
2770..2775  
/gene="FUT8"  
/note="putative"  
2869..2874  
/gene="FUT8"  
/note="putative"  
BASE COUNT 905 a 597 c 663 g 827 t





```
/product="glycoprotein 6-alpha-L-fucosyltransferase"
/protein_id="CAA76987.1"
/db_xref="GI:3451267"
/translaton="MRPTGSRWIMLILFWGTHLLFYIGHLVRDNDHPDHSRELS
KILAKLERKQONEDLRMAESLRIPEPIDOGPAIGRVRVLESQLVKAEQIENYKK
QTRNGLKHQILRRIRRENGAKELWFFLOSELKKNLEGNELQRHADDEFLLDGHHE
RSITMDLYLSQTCAGDWREKEAKDLTELVRITLYLQNPDKCSKAKLVNINKGC
GYGCLHHVYCFMAYCTQRTLLILESONWRYATGWETVFRPVSETCTDRSGISTGH
WSGEKDNQVOVELPIVDSLHPRPPLPLAVPDLADLVRVHGDPAVWVVSOFVKY
LIRPOWLEKIEBEATKGLGKFPVIGVHVPRTDKVTGEAAPHLEIEMVHVEHFQL
LARMQDKKRVILATDPSLLKKAETKIPNTEFISNLSWSAGLHNRFTENSILRGV
ILDHFLSQADFLVCTGQVCRVAYEIMQTLHPDASANFSLDDIYFYGQNAHNI
AIYAHOPRTADEIPMEPGDIIICVAGNHDWYSGVNRKLGRTGLYPSYKVRKEIETVK
YPTVPEAEK"
variation 2506
/ene="FUT8"
/ene="I"
/requence=".5"
/replace="c"
3046..3051
/ene="FUT8"
/ene="putative"
3058..3063
/ene="FUT8"
/ene="putative"
3157..3162
/ene="FUT8"
/ene="putative"
BASE COUNT 910 a 710 c 721 g 840 t 5 others
ORIGIN
Query Match 86.9%; Score 1501.4; DB 9; Length 3186;
Best Local Similarity 91.8%; Pred. No. 0;
Matches 1586; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 1 ATCGGGGATGACGTGGTTCCTGGCGTGGATATGCTCAITCTTTTTCCTGGGGGACC 60
DB 717 ATCGGGCCATGACGTGGTTCCTGGCGTGGATATGCTCAITCTTTTTCCTGGGGGACC 776
QY 61 TTGTTATTTTATATAGTGGTCAITTTGGTTCAGATATGACCCATCCATGATCCAGC 120
DB 777 TTGCTGTTTATATAGTGGTCAITTTGGTTCAGATATGACCCATCCATGATCCAGC 836
QY 121 AGAGACCTCCAGATCTTTCGAAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGA 180
DB 837 CGAGAACTGTCCAGATCTTTCGAAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGA 896
QY 181 AGGCGAATGGCTGAGTCTCTCCGAATACAGAGGCGCCCATTTGACGGGGACAGCTACA 240
DB 897 AGGCGAATGGCGAATCTCTCCGAATACAGAGGCGCCCATTTGATCAGGGGCCAGCTATA 956
QY 241 GGAAGAGTCCGTGTTTGAAGAACAGCTTGTGAAGGCCCAAGAACAGATGAAATTAAC 300
DB 957 GGAAGAGTACGCGTTTGAAGAACAGCTTGTGAAGGCCCAAGAACAGATGAAATTAAC 1016
QY 301 AAGAAACAGCTAGAAATGGTCTGGGAAGGATCATGAATCTTAAGAGGAGGATTCGA 360
DB 1017 AAGAAACAGACAGAAATGGTCTGGGAAGGATCATGAATCTTGAAGGAGGAGGATTCGA 1076
QY 361 AATGAGCTTAAGAGCTCTGGTGTTCCTCAAAAGGAACTGAAGAAATTAAGCATTTA 420
DB 1077 AATGAGCTTAAGAGCTCTGGTGTTCCTCAAAAGGAACTGAAGAAATTAAGCACTTA 1136
QY 421 GAAGAAATGAATCCAAAGACATGCAGATGAAATCTTTTGGATTTAGGACACCATGAA 480
DB 1137 GAAGAAATGAATCCAAAGACATGCAGATGAAATCTTTTGGATTTAGGACATCATGAA 1196
QY 481 AGGTCTATCATGACAGATCTATATCTTACCTCAGTCAAAACAGATGAGAGGAGGATGGCGT 540
DB 1197 AGGTCTATATATGACGAGTCTATATCTTACCTCAGTCAAAACAGATGAGAGGAGGATGGCGG 1256
QY 541 GAAAGAGGCGCAAGATCTGACAGAGCTGGTCCAGCGGAGATTAACATATCTCCAGAT 600
```

```
Db 1257 GAAAAGAGGCGCAAGAGATCTGACAGAACTGGTTCAGCGGAGAAATAATATCTTCAGAAAT 1316
QY 601 CCTAAGGACTGACGCA AAGCAGGAGAGCTGGTGTGAACATCAATAAAGGCTGTGGCTAT 660
DB 1317 CCCAAGGACTGACGCA AAGCAGGAGAGCTGGTGTGAATATCAACAAGGCTGTGGCTAT 1376
QY 661 GGTGTGCAACTCCATCAGCTGGTGTCTACTGTTTCATGATTCGTATATGGCACCCAGCAACA 720
DB 1377 GGCTGTGAGCTCCATCATGTTGGTCTCTACTGTTTCATGATTCGTATATGGCACCCAGCAACA 1436
QY 721 CTATCTTGGAACTCAGAAATGGCGCTATGCTACTGTTGGTGGATGGGAGACTGTGTATAGA 780
DB 1437 CTATCTTGGAACTCAGAAATGGCGCTATGCTACTGTTGGTGGATGGGAGACTGTATTTAGG 1496
QY 781 CCTGTAAGTGAGACATCTACAGACAGATCTGGCGCTCTCCACTGACACACTGCTCAGGTCAA 840
DB 1497 CCTGTAAGTGAGACATCTACAGACAGATCTGGCGCTCTCCACTGACACACTGCTCAGGTCAA 1556
QY 841 GTAAATGACAAAAACA TTTCAAGTGGTGGAGCTCCCAATTTGTAGACAGCTCCATCTCTCGG 900
DB 1557 GTGAAGGACAAAAATG TTTCAAGTGGTGGAGCTCCCAATTTGTAGACAGCTCTTCATCCCGGT 1616
QY 901 CCTCCTTACTTACCACTGGCTGTTCCAGAGACACTTGGAGACCGACTCTTAAGAGTCCAT 960
DB 1617 CCTCCATATTTACCCCTTGGCTGTACAGAGACACTTCGACAGACTTGTACGAGTGCAT 1676
QY 961 GGTGACCTCGAGTGTGGTGGTCTCCAGTTCTCAAACTACTTGTCTGTCACCAACT 1020
DB 1677 GGTGACCTCGAGTGTGGTGGTCTCTCAGTTTGTCAAACTACTTGTCTGTCGCGCCACAGCT 1736
QY 1021 TGGCTGAAAAGAAAAI TAGAAGAGACCAACAAGAGCTTGGCTTCAAACTCCAGTTATT 1080
DB 1737 TGGCTGAAAAGAAAAI TAGAAGAGACCAACAAGAGCTTGGCTTCAAACTCCAGTTATT 1796
QY 1081 GGAGTCCATGTACAGACACAGACAAAGTGGGAAACAGAGAGCTTCCACCCCATCCAG 1140
DB 1797 GGAGTCCATGTACAGACACAGACAAAGTGGGAAACAGAGAGCTTCCACCCCATCCAG 1856
QY 1141 GAGTACATGTACAGTGTGAAGACATTTTCAGCTTCTCCGACAGAGTCCAGATTCGAT 1200
DB 1857 GAGTACATGTGTCATGTTGAAGAACATTTTCAGCTTCTTCGACGAGAGTCCAGATTCGAC 1916
QY 1201 AAAAAAGAGTATATCTGGCTACTGATGATCTCTACTTTGTTAAAGAGGAGGCAAGCAAG 1260
DB 1917 AAAAAAGAGTATTTTGGCCACAGATGACCCCTTCTTTTAAAGAGGAGGCAAGCAAG 1976
QY 1261 TACTCCAATATGAATTTATTTAGTGATACTCTATTTTCTTGGTTCAGCTGGGACTACAAAT 1320
DB 1977 TACCCCAATATGAATTTATTTAGTGATACTCTATTTTCTTGGTTCAGCTGGGACTACAAAT 2036
QY 1321 CGGTACACAGAAATTTCACTTCGGGGTGTGATCTCTGATATACACTTTCTCTCAGAGCT 1380
DB 2037 CGGTACACAGAAATTTCACTTCGGGGTGTGATCTCTGATATACACTTTCTCTCAGAGCT 2096
QY 1381 GACTTCTAGTGTCTACTTTTTCATCCAGGTCTGTGGGTTGCTTATGAATCATGCAA 1440
DB 2097 GACTTCTAGTGTGTACTTTTTCATCCAGGTCTGTGGGTTGCTTATGAATCATGCAA 2156
QY 1441 ACCCTGCTATCTGATGCTCTCGAACTTCATTTCTTTGGTGTGATCATCTACTATTTTGG 1500
DB 2157 ACCTACATCTGATGCTCTCGAACTTCATTTCTTTTAGATGATCATCTACTATTTTGG 2216
QY 1501 GGCCAAATGGCCACATCAGATTTGCTTTTATCTCTCAAACTCCAGACTGAAGAGAA 1560
DB 2217 GGCCAGAAATGGCCACATCAGATTTGCTTTTATGCTCACCACCCCGCACTTCAGATGAA 2276
QY 1561 ATTCCAATGGAACTGGAGATCATTTGGTGTGGCTGAAACCATTTGGGATGGTATTCT 1620
DB 2277 ATTCCAATGGAACTGGAGATCATTTGGTGTGGCTGAAACCATTTGGGATGGTATTCT 2336
QY 1621 AAAGGTATCAACAGAAACTTTGGAAACAGAGCTTTATATCTCTCTCAAAAGTCCGAGAG 1680
```



Db 2337 AAGGTGTCACAGAAATTGGGAAGGCGCCTATATCCCTCTCAAGAAGTTCGAGAG 2399

Qy 1681 AAGATAGAAAACAGTCAAGTAGTATCCACATATATCCTCAAGCTGAAAAATA 1727  
|||||  
Db 2397 AAGATAGAACGGTCAAGTAGTATCCACATATCCTGAGGCTGAGAATA 2443  
|||||

RESULT 14  
HSY17979 3280 bp mRNA linear PRI 06-NOV-1998  
LOCUS Homo sapiens mRNA for glycoprotein 6-alpha-L-fucosyltransferase  
DEFINITION transcript A2.  
ACCESSION Y17979  
VERSION Y17979.1 GI:3451268  
KEYWORDS FUT8 gene; glycoprotein 6-alpha-L-fucosyltransferase.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1  
AUTHORS Cailleau,A., Balanzino,L., Candelier,J.J., Oriol,R. and  
Mellicone,R.  
TITLE Differential splice variants of human FUT8 embryonic cDNA  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3280)  
AUTHORS Cailleau,A.L.V.  
TITLE Direct Submission  
JOURNAL Submitted (19-AUG-1998) A.L.V. Cailleau, INSERM U178, 16 av P.V.  
COMMENT Couturier, 94807 Villejuif Cedex, FRANCE  
FEATURES Location/Qualifiers  
source 1..3280  
organism="Homo sapiens"  
mol\_type="mRNA"  
db\_xref="taxon:9606"  
map="14q23"  
clones="FUT8 transcript A2"  
tissue\_type="whole embryo"  
dev\_stage="40-70 days"  
gene 1..3280  
/gene="FUT8"  
exon 1..389  
/gene="FUT8"  
note="A"  
exon 390..488  
/gene="FUT8"  
note="C"  
exon 489..52444  
/gene="FUT8"  
note="D"  
CDS 717..2444  
/gene="FUT8"  
EC number="2.4.1.68"  
note="allele I"  
codon\_start=1  
product="glycoprotein 6-alpha-L-fucosyltransferase"  
protein\_id="CAA76988.1"  
db\_xref="GI:3451269"  
translation="MRPWTGSWRIMLILFAWGTLFLFYIGHLVRDNDDPHSHRSRELS  
KILAKRLRQONEDLRMAESLRIPEGPIDQGPAICGRVLEQLVKAEQIENYKK  
QTNRGLKHDEILLRRLENGAKELWFFLOSELKKLNELGRHADEFLLDLGHKE  
RSINTDIYLSTDGAGREKEADLTFLVQRRTYQLNPKDCSKAKVLVCNKIKG  
GYGQLHHVVCFMIAYGTORTILESONRYATGGWTFVRPVSETCTDRSGISTGH  
WSGVCKNQVQVBLPTVDLSLHPRPYLPLAVPEDLADRLRVRHGDPAVMVVSQFYKL  
LIRPOVLEKEIEBATKLGPGHPVIQGVHVRTDKVGTEAFAPHIEETMVHVEHFOL  
LARMOVDKKRXYLATDDPSLLKEAKTKYNBYEISDNSISWSAGLHNRYTENS LRGV  
ILDHIFLSQADELVCTFSOVCRAVEIMOTLHPDASNFSLDDIYYFGQNANHQOI  
AIYAQPRTADEIPNEPFDIIGVAGNHWDGYSKGVKNRKLRGTGLYPYSKYKREKITETVK  
YTPYPEAKN"  
variation 2506  
/gene="FUT8"



[illegible]

Search completed: February 2, 2004, 16:22:55  
Job time : 6350.17 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: February 2, 2004, 10:17:23 ; Search time 467.615 Seconds  
(without alignments)  
9975.371 Million cell updates/sec

Title: US-09-971-773-2  
Perfect score: 1728  
Sequence: 1 atgcggggcagctgcttc.....atcctgaagctgaaaaatag 1728

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues  
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq 19Jun03:\*

1:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1728	100.0	1728	24	ASK70032
2	1584	91.7	2008	24	ASK70031
3	1501.4	86.9	1759	24	ABR98809
4	1501.4	86.9	2100	18	AAT76574
5	1501.4	86.9	3291	23	ABV22468
6	1501.4	86.9	3291	23	ABV28283
7	1490	86.2	3007	22	AAH33315
8	1469.4	85.0	1728	18	AAT76573
					Antibody productio
					Antibody productio
					Alpha1,6-fucosyl t
					Human alpha 1-6 fu
					Human prostate exp
					Human prostate exp
					Human colon cancer
					Pig alpha 1-6 fuco

9	931	53.9	979	31	AAC63892	Rat FUT8 cDNA. Ra
10	931	53.9	979	34	ABK70037	Antibody productio
11	894.2	51.7	979	31	AAC63891	Chinese hamster FU
12	894.2	51.7	979	34	ABK70036	Antibody productio
13	876.2	50.7	1017	32	ABF87952	Human alpha 1-6 fu
14	864.6	35.0	699	32	AAF87953	Human alpha 1-6 fu
15	378.8	21.9	2761	33	ABLO4601	Drosophila melanog
16	236.4	17.2	394	20	AAV89049	EST clone CS100.
17	269.2	15.6	503	32	AAK11265	Human brain expres
18	269.2	15.6	503	32	AAI42879	Human brain expres
19	241.8	14.0	4682	33	ABLO4600	Drosophila melanog
20	210.8	12.2	551	32	AAK11037	Human brain expres
21	210.8	12.2	551	32	AAI42659	Probe #11345 used
22	205.4	11.9	248	32	ABA47883	Human foetal cell
23	205.4	11.9	248	32	ABA65776	Human foetal liver
24	205.4	11.9	248	32	ABA32861	Probe #11327 for g
25	205.4	11.9	248	32	AAK14177	Human brain expres
26	205.4	11.9	248	32	AAK25142	Human brain expres
27	205.4	11.9	248	32	AAK39911	Human bone marrow
28	205.4	11.9	248	32	AAI20722	Probe #10655 for g
29	205.4	11.9	248	32	AAI45939	Probe #14625 used
30	205.4	11.9	248	32	AAI57183	Probe #25869 used
31	205.4	11.9	248	32	AAI06420	Probe #6411 used t
32	205.4	11.9	248	32	ABS39501	Human liver single
33	205.4	11.9	248	32	ABS14010	Human genome-deriv
34	188.6	10.9	9196	24	ABK70033	Antibody productio
35	147.4	8.5	481	25	ABT21901	Breast cancer mark
36	126.8	7.3	384	22	ABA42755	Human breast cell
37	126.8	7.3	384	22	ABA53183	Human foetal liver
38	126.8	7.3	384	22	ABA22955	Probe #1421 for ge
39	126.8	7.3	384	22	AAK01429	Human brain expres
40	126.8	7.3	384	22	AAK26877	Human bone marrow
41	126.8	7.3	384	22	AAI11508	Probe #1441 for ge
42	126.8	7.3	384	22	AAI32785	Probe #1471 used t
43	126.8	7.3	384	22	AAI01422	Probe #1413 used t
44	126.8	7.3	384	23	ABS26469	Human liver single
45	126.8	7.3	384	24	ABS01477	Human genome-deriv

ALIGNMENTS

RESULT 1

ABK70032  
ID ABK70032 standard; cDNA; 1728 BP.  
AC ABK70032;  
XX 15-JUL-2002 (first entry)  
DT Antibody production method related cDNA #2.  
DE Antibody production; cytostatic; immunomodulator; vasotropic; virucide;  
XX Antibacterial; antiinflammatory; antiallergic; allergy; inflammation;  
KW autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO;  
KW tumour; circulatory disease; infection; primer; ss.  
XX Mus musculus.  
XX WO200231140-A1.  
XX 18-APR-2002.  
XX 05-OCT-2001; 2001WO-JP08804.  
XX 06-OCT-2000; 2000JP-0308526.  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX Kanda Y, Satoh M, Nakamura K, Uchida K, Shinkawa T, Yamane N;  
XX Hosaka E, Yamano K, Yamasaki M, Hanai N;  
XX WPI, 2002-340182/37.



ABK70031  
ID ABK70031 standard; cDNA; 2008 BP.  
XX AC ABK70031;  
XX  
DT 15-JUL-2002 (first entry)  
XX  
DE DE  
XX  
XX  
XX  
KW Antibody production method related cdna #1.  
KW Antibody production; cytostatic; immunomodulator; vasotropic; virucide;  
KW antibacterial; antiinflammatory; antiallergic; allergy; inflammation;  
KW autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO;  
KW tumour; circulatory disease; infection; primer; ss.  
XX  
OS Cricetus griseus.  
XX  
XX WO200231140-A1.  
XX  
XX 18-APR-2002.  
XX  
XX 05-OCT-2001; 2001WO-JP08804.  
XX  
XX 06-OCT-2000; 2000JP-0308526.  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
XX Kanda Y, Satoh M, Nakamura K, Uchida K, Shinkawa T, Yamane N;  
XX Hosaka E, Yamano K, Yamasaki M, Hanai N;  
XX  
XX WPI; 2002-340182/37.  
XX  
XX  
XX Cells producing antibody compositions including antibody fragments and  
XX fusion proteins with Fc domain of antibody, useful for prevention or  
XX treatment of cancer, immune diseases, circulatory diseases and  
XX infections  
XX  
XX Claim 13; Page 1-3; 314pp; Japanese.  
XX  
XX This invention relates to novel method for antibody production  
XX comprising a Chinese hamster ovarian tissue-originated (CHO) cell  
XX transfected with a gene encoding an antibody molecule for producing a  
XX composition comprising an antibody molecule with an Fc domain bonded  
XX to the N-glycoside linkage complex sugar chain. The produced antibody  
XX compositions are drugs for prevention or treatment of diseases  
XX accompanying tumour, allergy or inflammation, autoimmune diseases,  
XX circulatory diseases, and viral and bacterial infections. The  
XX antibodies can be stably produced using the method of the invention  
XX with high binding activity and potency thus leading to high safety and  
XX reduced side effects when applied alone or in combination with other  
XX drugs for therapy. The present sequence represents a nucleotide  
XX molecule used in the method of the invention.  
XX  
XX Sequence 2008 BP; 626 A; 425 C; 465 G; 492 T; 0 other;  
XX  
XX Query Match 91.7%; Score 1584; DB 24; Length 2008;  
XX Best Local Similarity 94.8%; Pred No. 0;  
XX Matches 1638; Conservative 0; Mismatches 90; Indels 0; Gaps 0;  
XX  
XX 1 ATGCGGGCATGGACTGGTTCCTGGCGTTGGATTATGCTCAATCTTTTGGCTGGGGACC 60  
XX  
XX 100 ATGCGGGCATGGACTGGTTCCTGGCGTTGGATTATGCTCAATCTTTTGGCTGGGGACC 159  
XX  
XX 61 TTGTTATTTTATAGTGGTTCATTTGGTTCGAGATATGACCCTTGATCATCTCCAGC 120  
XX  
XX 160 TTATTTGTTTATATAGTGGTTCATTTGGTTCGAGATATGACCCTTGATCATCTCCAGC 219  
XX  
XX 121 AGAGAACTCTCAAGATTCTTGAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 180  
XX  
XX 220 AGAGAACTCTCAAGATTCTTGAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 279  
XX  
XX 181 AGGCGAATGGCTGAGTCTCTCCGAATACCAAGAGCCCATTTGACAGGGGACAGCTACA 240  
XX  
XX 280 AGGAGATGGCTGAGTCTCTCCGAATACCAAGAGCCCATTTGATCAGGGGACAGCTACA 339  
XX

241 GGAAGAGTCCGTGTTTAAAGAAACAGCTTGTAAAGGCCAAAGAACAGATTGAAAATTAC 300  
340 GGAAGAGTCCGTGTTTAAAGAAACAGCTTGTAAAGGCCAAAGAACAGATTGAAAATTAC 399  
301 AAGAAACAAGCTAGAAATGCTCGGGAAGAGTATGAAATCTTAAGAGGAGGATTGAA 360  
400 AAGAAACAAGCTAGAAATGCTCGGGAAGAGTATGAAATCTTAAGAGGAGGATTGAA 459  
361 AATGAGCTAAAGAGCTCTGGTCTTCTCAAAAGCGAATCTGAAGAAATTAAGCATTTA 420  
460 AATGAGCTAAAGAGCTCTGGTCTTCTCAAAAGCGAATCTGAAGAAATTAAGAAATTA 519  
421 GAAGGAAATGAATCTCAAAAGCATGCGAGTGAATCTTTTGGATTTAGGACATCATGAA 480  
520 GAAGGAAACGAATCTCAAAAGCATGCGAGTGAATCTTTTGGATTTAGGACATCATGAA 579  
481 AGGTCTATGACAGATCTATCTACTAGTCAACAGATGAGAGCGGGATTTGGCT 540  
580 AGGTCTATGACAGATCTATCTACTAGTCAACAGATGAGAGCGGGATTTGGCT 639  
541 GAAAAGAGGCGCAAGATCTGACAGAGCTGTCAGGCGGAGAAATAACATATCTCCAGAA 600  
640 GAAAAGAGGCGCAAGATCTGACAGAGCTGTCAGGCGGAGAAATAACATATCTCCAGAA 699  
601 CCTAAGGACTGACAGCAAGCCAGGAAGCTGCTGTGTAAACATCAATAAAGGCTGTGGCTAT 660  
700 CCTAAGGACTGACAGCAAGCCAGGAAGCTGCTGTGTAAATAATCAACAAAGGCTGTGGCTAT 759  
661 GGTTGTCAATCTCAATCAGTGGTCTACTGTTTCTATGATTTGTCAGTGGCACCAGCGGAA 720  
760 GGATGTCAATCTCAATCAGTGGTCTACTGTTTCTATGATTTGTCAGTGGCACCAGCGGAA 819  
721 CTCAATCTTGAATCTCAAGATTTGGCGTATCTCTACTGCTGATGGAGACTGTGTTAGA 780  
820 CTCAATCTTGAATCTCAAGATTTGGCGTATCTCTACTGCTGATGGAGACTGTGTTAGA 879  
781 CCTGTAAGTGAGACATGTACAGACAGATCTGGCTCTCCACTGGACACTGGTCCAGTGAA 840  
880 CCTGTAAGTGAGACATGTACAGACAGATCTGGCTCTCCACTGGACACTGGTCCAGTGAA 939  
841 GTAAATGACAAAAACATTCAGAGTGGTGGCTGCTCCCATTTGTAGACAGCTCCATCCTCG 900  
940 GTGAAGACAAAAATGTTCAAGTGGTGGCTGCTCCCATTTGTAGACAGCTCCATCCTCGT 999  
901 CCTCTTACTTACCACTGGCTGTTCCAGAAACCTTGCAGACCGACTCCTTAAGAGTCCAT 960  
1000 CCTCTTACTTACCACTGGCTGTTCCAGAAACCTTGCAGACCGACTCCTTAAGAGTCCAT 1059  
961 GGTGACCTGTCAGTGTGTGGGTGTCCAGATTTGTCAAAATCTTGTATTCGTTCCACAACT 1020  
1060 GGTGATCTCTGAGTGTGTGGGTATCCAGTTTGTCAATATCTTGTATCCGTTCCACAACT 1119  
1021 TGGCTGGAAGAAAGAAATAGAAAGCCACCAAGAGCTTGGCTTCAAAATCTCAAGTATTT 1080  
1120 TGGCTGGAAGAAAGAAATAGAAAGCCACCAAGAGCTTGGCTTCAAAATCTCAAGTATTT 1179  
1081 GGAGTCCATGTTCAGACGACAGCAAAAGTGGGAAAGAGAGCTTCCACCCCATCCAG 1140  
1180 GGAGTCCATGTTCAGACGACCTGACAAAGTGGGAAAGAGAGCTTCCATCCCATCCAG 1239  
1141 GAGTACATGTTGACAGTTGAAGAAACATTTTCAGCTTCTCGCAACGAGAAATGCAAGTGGAT 1200  
1240 GAATACATGTTGACAGTTGAAGAAACATTTTCAGCTTCTCGCAACGAGAAATGCAAGTGGAT 1299  
1201 AAAAAAGAGTATATCTGGCTACTGATGATCTTCTTGTAAAGAGGAGGCAAGCAAG 1260  
1300 AAAAAAGAGTATATCTGGCTACTGATGATCTTCTTGTAAAGAGGAGGCAAGCAAG 1359  
1261 TACTCCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1320  
1360 TACTCCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1419

```
QY 1321 CGGTACACAGAAATTCACCTCGGGGTGATCCTGGATATACACTTTCTCTCAGGCT 1380
|||
DB 1420 CGATACACAGAAATTCACCTCGGGCGGTGATCCTGGATATACACTTTCTCTCCAGCT 1479
|||
QY 1381 GACTTCTTAGTGTGTAATTTTTCATCCAGGCTGCTCGGGTGTGCTTATGAATCATGCAA 1440
|||
DB 1480 GACTTCTTAGTGTGTAATTTTTCATCCAGGCTGCTCGGGTGTGCTTATGAATCATGCAA 1539
|||
QY 1441 ACCCTGCATCTGATCGCTCTCGCAACTTCCATTTCTTTGGATGACATCTACTATTTTGA 1500
|||
DB 1540 ACATCTGCATCTGATCGCTCTCGCAACTTCCATTTCTTTAGATGACATCTACTATTTTGA 1599
|||
QY 1501 GGCCTAAATGCCCAATCAGATTGCTGTTTATCTCTCAAACTCGAACTGAAGAGAA 1560
|||
DB 1600 GGCCTAAATGCCCAATCAGATTGCTGTTTATCTCTCAAACTCGAACTGAAGAGAA 1659
|||
QY 1561 ATTCCAAATGGAACCTGGAGATATCATTTGTTGGCTGGAACCATTTGGATGTTTATCT 1620
|||
DB 1660 ATCCCAATGGAACCTGGAGATATCATTTGTTGGCTGGAACCATTTGGATGTTTATCT 1719
|||
QY 1621 AAAGGTATCAACAGAAAACTTGGAAAAACAGGCTTTATATCCCTCTCAAAAGTCCGAGAG 1680
|||
DB 1720 AAAGGTGTCAACAGAAAACTAGAAAAACAGGCTTTATATCCCTCTCAAAAGTCCGAGAG 1779
|||
QY 1681 AGATAGAAACAGTCAAGTATCCCATATATCTGAACTGAAAAATAG 1728
|||
DB 1780 AGATAGAAACAGTCAAAATACCTCATATATCTGAACTGAAAAATAG 1827
|||
```

## RESULT 3

```
ABA98809
ID ABA98809 standard; cDNA; 1759 BP.
XX AC ABA98809;
XX DT 07-MAY-2002 (first entry)
XX DE Alphas,6-fucosyl transferase cDNA.
XX KW Plant; glycoprotein; alphas,6-fucosyl transferase; alphas,6-PT;
XX KW enzyme; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 17..1744
XX FT /*tag= a
XX FT /product= "Alphas,6-fucosyl transferase"
XX PN JP2001333787-A.
XX PD 04-DEC-2001.
XX PF 06-MAR-2001; 2001JP-0062704.
XX PR 22-MAR-2000; 2000JP-0081059.
XX PA (TANI/) TANIGUCHI N.
XX PA (SEKI/) SEKI T.
XX PA (FUJI/) FUJIYAMA K.
XX DR WPI; 2002-158816/21.
XX DR P-PSDB; ABB08405.
XX PT A plant cell with an animal type sugar chain adding function, for the
XX PT preparation of a glycoprotein with an animal type sugar chain -
XX PS Example 1; Page 22-24; 38pp; Japanese.
XX CC The invention relates to a plant cell with an animal type sugar chain
XX CC adding function, created by transforming a tobacco cell with a gene
XX CC encoding an enzyme derived from an animal which can transfer a fucose
XX CC residue to the reductive end acetylglucosamine residue of a sugar chain.
```

```
CC The gene that is introduced into the plant cell encodes the enzyme
CC alphas,6-fucosyl transferase. The method of the invention is useful for
CC the preparation of a glycoprotein having animal type sugar chain. The
CC current sequence represents alphas,6-fucosyl transferase cDNA.
```

XX Sequence 1759 BP; 533 A; 367 C; 418 G; 441 T; 0 other;

Query Match 86.9%; Score 1501.4; DB 24; Length 1759;  
Best Local Similarity 91.8%; Pred. No. 0;  
Matches 1586; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```
QY 1 ATCGGGGCATCGACTGGTTCCTGGCTTGGATTATGCTCATTTCTTTTGGCTGGGGACC 60
|||
DB 17 ATCGGGGCATCGACTGGTTCCTGGCTTGGATTATGCTCATTTCTTTTGGCTGGGGACC 76
|||
QY 61 TTGTTATTTTATATAGTGGTTCATTTGTTTCGAGATATGACCACTTCATCTCAGC 120
|||
DB 77 TTGCTGTTTTATATAGTGGTTCATTTGTTTCGAGATATGACCACTTCATCTCAGC 136
|||
QY 121 AGAGAACTCTCAAGATTTCTGCAAGCTTGAACGCTTAAACAGCAAAATGAAGCTTG 180
|||
DB 137 CGAGAACTGTCTCAAGATTTCTGCAAGCTTGAACGCTTAAACAGCAAAATGAAGCTTG 196
|||
QY 181 AGCGAAATGGCTGAGTCTCTCCGAATACCAAGGCCCTTATGACAGGGGACAGCTACA 240
|||
DB 197 AGCGAAATGGCGGAAATCTCTCCGAATACCAAGGCCCTTATGATCAGGGGCCAGCTATA 256
|||
QY 241 GGAAGAGTCCGTGTTTGTAGAAAGACAGCTTCTTAAAGCCAAAGAACAGATGAAATATAC 300
|||
DB 257 GGAAGAGTACGCGTTTTAGAAAGACAGCTTCTTAAAGCCAAAGAACAGATGAAATATAC 316
|||
QY 301 AAGAAACAGCTAGAAATGCTCTGGGAGAGATCATGAAATCTTAAAGAGGAGGATGAA 360
|||
DB 317 AAGAAACAGACAGAAATGCTCTGGGAGAGATCATGAAATCTTAAAGAGGAGGATGAA 376
|||
QY 361 AATGAGCTAAAGAGCTCTGTTTTCTCAAGAGCGAACTGAAGAAATTTAAAGCATTTA 420
|||
DB 377 AATGAGCTAAAGAGCTCTGTTTTCTCAAGAGTGAATTTGAAGAAATTTAAAGAACTTA 436
|||
QY 421 GAAGAAATGAATCCCAAGACATGCAGATGAAATTTCTTTGGATTTAGGACACCATGAA 480
|||
DB 437 GAAGAAATGAATCCCAAGACATGCAGATGAAATTTCTTTGGATTTAGGACATCATGAA 496
|||
QY 481 AGGCTATATGACAGATCTATATCTCAGTCAACAGATGAGCAGGGGATTTGGGCT 540
|||
DB 497 AGGCTATATGACAGGATCTATATCTCAGTCAACAGATGAGCAGGGGATTTGGGCG 556
|||
QY 541 GAAAAGAGGCCAAAGATCTGACAGAGCTGCTCAGCGGAGAAATAACATATCTCCAGAA 600
|||
DB 557 GAAAAGAGGCCAAAGATCTGACAGAACTGGTTCAAGCGAGAAATAACATATCTCCAGAA 616
|||
QY 601 CCTAAGGACTGACAGCAAGCCAGGAAGCTGGTGTGTAACATCAATAAAGGCTGTGGCTAT 660
|||
DB 617 CCCAAGGACTGACAGCAAGCCAAAGCTGGTGTGTAATAACAACAAAGGCTGTGGCTAT 676
|||
QY 661 GGTGTCAACTCCATCAGCTGCTCTATGTTTTCATGATTGCTTATGGCACCCAGGAAACA 720
|||
DB 677 GGTGTCAAGTCCATCATGTGCTCTACTGCTTCATGATTGCAATATGGCACCCAGGAAACA 736
|||
QY 721 CTCATCTTTGGAATCTCAGAAATTTGGCGCTATGCTACTGTTGGATGGAGACTGTGTTTGA 780
|||
DB 737 CTCATCTTTGGAATCTCAGAAATTTGGCGCTATGCTACTGTTGGATGGAGACTGTATTTAG 796
|||
QY 781 CCTGTAGTGAGACATGTACAGACAGATCTCGGCTCTCCACTGGACACTGGTCAGGTGAA 840
|||
DB 797 CCTGTAGTGAGACATGTACAGACAGATCTCGGCTCTCCACTGGACACTGGTCAGGTGAA 856
|||
QY 841 GTAAATGACAAAAACATTTCAAGTGTTCAGCTCCCCATTTCTAGACAGCTCCATCTCGG 900
|||
DB 857 GTGAAGGACAAAAATGTTTCAAGTGTTCAGCTTCCATTTGTAGACAGTCTTCTATCCCCGT 916
|||
QY 901 CCTCTTACTTACCACTGGCTGTTTCCAGAAAGACCTTTGCAGACCCGACTCTTAAGAGTCCAT 960
|||
```



Db 917 CCTCCATATTACCCTGGCTGTACCAAGAGACCTCGCAGATCGACTTGTACGAGTGCAT 976  
QY GGTGACCTCGAGTGTGGTGTCCAGTTTGTCAATATCTGATTCGTCACCAACT 1020  
Db 977 GGTGACCTCGAGTGTGGTGTCTCAGTTTGTCAATATCTGATTCGTCACCAACT 1036  
QY TGGCTGGAAAAAGGAATAGAAGACCCACCAAGAGCTTGGCTTCAACATCCAGTTATT 1080  
Db 1037 TGGCTAGAAAAAGGAATAGAAGACCCACCAAGAGCTTGGCTTCAACATCCAGTTATT 1096  
QY GGAGTCCATGTGACGACGACAGCAAAAGTGGAGCAGAGCTTCCACCCCATCGAG 1140  
Db 1097 GGAGTCCATGTGACGACGACAGCAAAAGTGGAGCAGAGCTTCCATCCCATGAA 1156  
QY 1141 GAGTACATGGTACAGTTTGAAGACATTTTTCAGCTTCTCGCAGCAGAAATGCAAGTGCAT 1200  
Db 1157 GAGTACATGGTGCATGTTGAAGACATTTTTCAGCTTCTCGCAGCAGAAATGCAAGTGCAT 1216  
QY 1201 AAAAAAGAGTATATCTGGCTACTGATGATCCTACTTTTAAAGGAGGCAAGCAAAAG 1260  
Db 1217 AAAAAAGAGTATATCTGGCTACTGATGATCCTACTTTTAAAGGAGGCAAGCAAAAG 1276  
QY 1261 TACTCCAAATATGAATTTATGATGATTAATCTATTTCTTGGTCTAGCTGGACTACAAAT 1320  
Db 1277 TACCCCAATATGAATTTATGATGATTAATCTATTTCTTGGTCTAGCTGGACTACAAAT 1336  
QY 1321 CGGTACACAGAAATTCACCTCGGGTGTGATCCTGGATATACATTTCTCTCACAGGCT 1380  
Db 1337 CGGTACACAGAAATTCACCTCGGGTGTGATCCTGGATATACATTTCTCTCTCAGGCA 1396  
QY 1381 GACTTCTAGTGTGATCTTTTTCATCCAGGCTCTGCGGTTGCTTATGAATCATGCAA 1440  
Db 1397 GACTTCTAGTGTGATCTTTTTCATCCAGGCTCTGCGGTTGCTTATGAATCATGCAA 1456  
QY 1441 ACCCTGCATCTGATGCTCTGCGAATTCATCTTTTGGATGATCACTATTTTGGG 1500  
Db 1457 ACACATACATCTGATGCTCTGCGAATTCATCTTTTGGATGATCACTATTTTGGG 1516  
QY 1501 GGCCAAATGCCCAATCAGATTGCTTTATCTCTCAACACCTCGAATCGAAGGAA 1560  
Db 1517 GGCCAAATGCCCAATCAGATTGCTTTATCTCTCAACACCTCGAATCGAAGGAA 1576  
QY 1561 ATTCCAAATGGAACTCGAGATATCATTTGTGTGGCTGGAAACCATTTGGATGGTTATTCT 1620  
Db 1577 ATTCCAAATGGAACTCGAGATATCATTTGTGTGGCTGGAAACCATTTGGATGGTTATTCT 1636  
QY 1621 AAGGTATCAACAGAAAACTTGGAAAAACAGGCTTATATCCCTCTCAAAAGTCCGAGAG 1680  
Db 1637 AAGGTATCAACAGAAAACTTGGAAAAACAGGCTTATATCCCTCTCAAAAGTCCGAGAG 1696  
QY 1681 AAGATAGAAACAGTCAAGTATCCCATATATCTGAGCTGAAAAATA 1727  
Db 1697 AAGATAGAAACAGTCAAGTATCCCATATATCTGAGCTGAAAAATA 1743

## RESULT 4

AAT76574

ID AAT76574 standard; cDNA to mRNA; 2100 BP.

XX AC

XX AAT76574;

AC AAT76574;

DT 05-MAR-1998 (first entry)

DE Human alpha 1-6 fucosyltransferase gene.

DE Human alpha 1-6 fucosyltransferase gene.

XX Alpha 1-6 fucosyltransferase; enzyme; pig; human; fucose transfer;

KW guanosine diphosphate; sugar chain synthesis; modification; antibody;

KW GlnAc; cancer diagnosis; ss.

XX Homo sapiens.

OS Homo sapiens.

XX Key

PH Location/Qualifiers

FT CDS

198..1925

/\*tag= a

W09727303-A1.

31-JUL-1997.

23-JAN-1997; 97WO-JP00171.

22-JUL-1996; 96JP-0192360.

24-JAN-1996; 96JP-0010365.

21-JUN-1996; 96JP-0161648.

24-JUN-1996; 96JP-0162813.

(TOYM) TOYO BOSEKI KK.

Shiba T, Taniguchi N, Uozumi N, Yanagidani S;

P-PSDB; AAW22125.

WPI; 1997-393690/36.

Human or pig alpha 1-6 fucosyltransferase and DNA encoding it - for

synthesis and modification of sugar chains and used as an antigen

for production of diagnostic antibodies

Claim 18; Page 39-43; 61pp; Japanese.

AAT76573 and AAT76574 represent the coding sequences for the pig and

human alpha 1-6 fucosyltransferases of the invention, respectively. The

6-hydroxyl group of the GlcNAc nearest to R in the receptor molecule:

GlcNAc 1-4GlcNAc-6 (GlcNAc 1-2Manalpa 1-3)Manbeta

1-4GlcNAc 1-4GlcNAc-6 (GlcNAc 1-2Manalpa 1-6)GlcNAc-R. It has

an optimum pH of about 7.0 (pig) or 7.5 (human), and is stable over the

pH range 4-10 after 5 hours at 4 degrees C. The optimum working

temperature of the enzyme is 30-37 degrees C. A bivalent metal is not

required for activity of the enzyme, and the enzyme is not inhibited in

the presence of 5 mM EDTA. The enzyme is useful in the synthesis and

modification of sugar chains, and as antigen for the production of

antibodies recognising the enzyme. The antibodies can be used for the

diagnosis of cancer and other diseases.

Sequence 2100 BP; 651 A; 436 C; 489 G; 524 T; 0 other;

Query Match 86.9%; Score 1501.4; DB 18; Length 2100;

Best Local Similarity 91.8%; Pred. No. 0;

Matches 1586; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 1 ATGCGGGCATGGACTGGTTCCTCGGTTGGATTATGCTCATTTTTCCTGGGGACC 60

198 ATGCGGGCATGGACTGGTTCCTCGGTTGGATTATGCTCATTTTTCCTGGGGACC 257

61 TTGTTATTTTATAGTGGTTCATTTGGTTCGAGATATGACCCCTGATCCTCCAGC 120

258 TTGCTGTTTATAGTGGTTCATTTGGTTCGAGATATGACCCCTGATCCTCCAGC 317

121 AGAGACTCTCCAGATTTCTGCAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 180

318 CGAGAACTGTCCAAGATTTCTGCAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 377

181 AGGCAATGGCTGAGTCTCTCCGAATACCAAGGCCCTTGTGACAGGGACAGCTACA 240

378 AGGCAATGGCTGAGTCTCTCCGAATACCAAGGCCCTTGTGACAGGGACAGCTACA 437

241 GGAAGAGTCCGTGTTTATAGAGAACAGCTTGTGTTAGGCCCAAGAACAGATTGAAATAC 300

438 GGAAGAGTCCGTGTTTATAGAGAACAGCTTGTGTTAGGCCCAAGAACAGATTGAAATAC 497

301 AAGAAACAGCTAGAAATGGTCTGGGAGAGCATGTAATCTTAAAGAGGAGGATTGAA 360

498 AAGAAACAGCTAGAAATGGTCTGGGAGAGCATGTAATCTTAAAGAGGAGGATTGAA 557

361 AATGAGAGTAAAGAGCTCTGGTCTTCTCTCAAAAGCAAGTGAAGAAATTAAGCATTTA 420



Db 558 AATGAGCTAAAGAGCTCTGGTTTTTCTTACAGAGTGAATTAAGAAACTTA 617  
Qy 421 GAAGGAATGAACTCAAGACATGAGATGAATTTCTTTGGATTTAGGACACCATGAA 480  
Db 618 GAAGGAATGAACTCAAGACATGAGATGAATTTCTTTGGATTTAGGACATGAA 677  
Qy 481 AGGTCTATCATGACAGATCTATACCTCAGTCAAAACAGATGGAGCGGATTTGGCGT 540  
Db 678 AGGTCTATAATGACGATCTATACCTCAGTCAAGACAGATGGAGCGGATTTGGCGG 737  
Qy 541 GAAAGAGGCCAAAGATCTGACAGCTGGTTCAGCGGAGATTAACATATCTCCAGAT 600  
Db 738 GAAAGAGGCCAAAGATCTGACAGATCTGGTTCAGCGGAGATTAACATATCTCCAGAT 797  
Qy 601 CCTAAGGACTGACGAAAGCCAGGAGCTGGTGTCTTAACATCAATAAAGGCTGTGGCTAT 660  
Db 798 CCCAAGGACTGACGAAAGCCAAAGCTGGTGTCTTAACATCAATAAAGGCTGTGGCTAT 857  
Qy 661 GGTTCTCAACTCCATCAGTGGTCTACTGTTTTCATGATTTGTCGACCCAGCGAACA 720  
Db 858 GGTCTGAGCTCCATCATGCTGTCTACTGCTTCATGATTTGTCATATGGCACCCAGCGACA 917  
Qy 721 CTCATCTTGGAAATCTCAGAAATTTGGCGCTATGCTACTGTTGATGGAGACTGTGTTAGA 780  
Db 918 CTCATCTTGGAAATCTCAGAAATTTGGCGCTATGCTACTGTTGATGGAGACTGTATTTAGG 977  
Qy 781 CCTGTAAGTGAGACATGTACAGACAGATCTGGCTCTCCACTGGACACTGTGTCAGTGCAA 840  
Db 978 CCTGTAAGTGAGACATGTACAGACAGATCTGGCTCTCCACTGGACACTGTGTCAGTGCAA 1037  
Qy 841 GTAAATGACAAAAATTCAGTGTGAGCTGCCATTTGTAGACAGCTCCATCTCTCGG 900  
Db 1038 GTGAGGACAAAAATTTCAAGTGTGAGCTGCCATTTGTAGACAGCTCTCATCCCCGT 1097  
Qy 901 CCTCTTACTTACCATGCTGCTTCCAGAGACCTTTGACAGCCGACTCTTAAGAGTCCAT 960  
Db 1098 CCTCCATATTTACCTTGGCTGTACAGAGACCTTCGACAGTGTGACAGTGTGACAGTGCAT 1157  
Qy 961 GGTGACCTGAGTGTGGTGTCCAGTTTGTCAATCTTCAATCTGTCACCAACT 1020  
Db 1158 GGTGACCTGAGTGTGGTGTCCAGTTTGTCAATCTTCAATCTGTCACCAACT 1217  
Qy 1021 TGGCTGGAAGAAATGAGAGCCACCAAGAGCTTGGCTTCAACATCCAGTTAT 1080  
Db 1218 TGGCTGGAAGAAATGAGAGCCACCAAGAGCTTGGCTTCAACATCCAGTTAT 1277  
Qy 1081 GGAGTCCATGTGACGACGACAGAAAGTGGGAAACAGAGCCCTTCCACCCCATTCGAG 1140  
Db 1278 GGAGTCCATGTGACGACGACAGAAAGTGGGAAACAGAGCTTCCATCCCATTCGAA 1337  
Qy 1141 GAGTACATGGTACAGTGTGAAGACATTTTCAGCTTCTCGACGAGAGTCAAGTGGAT 1200  
Db 1338 GAGTACATGGTGTGAGAAATTTTCAGCTTCTTCAGCTTCTTCAGCTGAGATGCAAGTGCAC 1397  
Qy 1201 AAAAAAGAGATATCTGCTACTCATGATCTTCTTTGTTAAAGAGGCGCAAGCAAG 1260  
Db 1398 AAAAAAGAGTATTTGGCCACAGATGACCTTCTTTAAGAGGCGCAAGCAAG 1457  
Qy 1261 TACTCCAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1320  
Db 1458 TACCCCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1517  
Qy 1321 CGGTACACAGAAATTTCACTTGGGCTGTGATCTGATATACATCTTCTCTCAGAGCT 1380  
Db 1518 CGATACACAGAAATTTCACTTGGGAGTGTGATCTGATATACATCTTCTCTCAGGCA 1577  
Qy 1381 GACTTCTTAGTGTACTTTTTTTCATCCAGGCTCTGCGGTTGCTTATGAATCATGCAA 1440  
Db 1578 GACTTCTTAGTGTACTTTTTTTCATCCAGGCTCTGCGGTTGCTTATGAATCATGCAA 1637  
Qy 1441 ACCCTGATCTGATGCTCTGCGAACTTCATTTCTTTGGATGACATCTACTATTTTGA 1500

Db 1638 ACATACATCTGATGCTCTGCAAACTTCTTCTTAGATGACATCTACTATTTTGGG 1697  
Qy 1501 GGCCAAAATGCCCAATCAGATTTGCTTTATCTCTCAAAACCTCGAACTGAAGAGGAA 1560  
Db 1698 GGCCAGATGCCCAATCAAAATTTGCCATTTATCTCACCACCAACCCGAACTGCAGATGAA 1757  
Qy 1561 ATTCCAATGGAACTGGAGATATCATTTGGTGTGCTGGAACCAATTTGGGATGTTATTTCT 1620  
Db 1758 ATTCCCATGGAACTGGAGATATCATTTGGTGTGCTGGAATCATTTGGGATGGCTATTTCT 1817  
Qy 1621 AAAGGTATCAACAGAAACTTTGGAACACAGGCTTATATCTCTCTCAAAAGTCCGAGAG 1680  
Db 1818 AAAGGTATCAACAGAAACTTTGGAAGGACGCGCTATATCTCTCTCAAAAGTTCGAGAG 1877  
Qy 1681 AAGATAGAAAACAGTCAAGATATCCACATATCTCTGAAGCTGAAAAATA 1727  
Db 1878 AAGATAGAAAACAGTCAAGATATCCACATATCTCTGAGCTGAAAAATA 1924  
RESULT 5  
ABV22468  
ID ABV22468 standard; cDNA; 3291 BP.  
XX AC ABV22468;  
XX DT 13-SEP-2002 (first entry)  
XX Human prostate expression marker cDNA 22459.  
DE Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX OS Homo sapiens.  
XX PN WO200160860-A2.  
XX PD 23-AUG-2001.  
XX PF 20-FEB-2001; 2001WO-US05171.  
XX PR 17-FEB-2000; 2000US-183319P.  
PR 16-MAR-2000; 2000US-189862P.  
PR 25-MAY-2000; 2000US-207454P.  
PR 09-JUN-2000; 2000US-211314P.  
PR 18-JUL-2000; 2000US-219007P.  
PR 13-DEC-2000; 2000US-255281P.  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
PA Schlegel R, Endege WO, Monahan JE;  
XX WPI; 2001-662795/76.  
XX Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -  
Claim 1; Page 3914; 11750pp; English.  
XX The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a patient;

CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

```
XX SQ Sequence 3291 BP; 925 A; 729 C; 745 G; 871 T; 21 other;
Query Match 86.9%; Score 1501.4; DB 23; Length 3291;
Best Local Similarity 91.8%; Pred. No. 0;
Matches 1586; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 1 ATGCGGGATGAGTGGTTCCTGGCGTTGGATTATGCTCATCTCTTTTGGCCCTGGGGGACC 60
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 727 ATGCGGCCATGAGTGGTTCCTGGCGTTGGATTATGCTCATCTCTTTTGGCCCTGGGGGACC 786
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 TTGTTATTTTATPATAGTGGTCAITTTGGTTCGAGATAATGACCAACCCCTGATCCTCCAGC 120
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 787 TTGCTGTTTTATATAGTGGTCAITTTGGTTCGAGATAATGACCAACCTGATCCTTAGC 846
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 AGAGAACTCTCAAGATTCTTCAAGCTTGAACCTTGAACCTTGAACCTTGAACCTTGAACCTTGA 180
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 847 CGAGAACTGTCCAGATTCTCGCAAGCTTGAACCTTGAACCTTGAACCTTGAACCTTGAACCTTGA 906
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 AGCGAAATGGCTGAGTCTCTCCGAATACCAAGAGCCCATTTGACCGGGGACAGCTACA 240
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 907 AGCGAAATGGCGAATCTCTCGGAATACCAAGAGCCCATTTGATCAGGGGCCAGCTATA 966
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 GGAAGAGTCCGCTGTTTTAGAGAAACAGCTTTGTTAAGGCCAAAGAACAGATTGAAAATTAC 300
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 967 GGAAGAGTACGCGTTTTAGAGAGCAGCTTTGTTAAGGCCAAAGAACAGATTGAAAATTAC 1026
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 301 AAGAAACAGCTAGAAATGGTCTGGGGAAGGATCATGAAATTTAAGAGAGGAGATTGAA 360
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1027 AAGAAACAGACAGAAATGGTCTGGGGAAGGATCATGAAATTTAAGAGAGGAGATTGAA 1086
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 361 AATGAGCTAAAGAGCTCTGTTTTTTCTACAAAGCGAACTGAAGAAATTTAAGAGATTTA 420
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1087 AATGAGCTAAAGAGCTCTGTTTTTTCTACAGAGTGAATTTAAGAAATTTAAGAACTTA 1146
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 421 GAAGAAATGAACTCCAAAGACATGACAGATGAAATTTCTTTGGATTTAGGACACCATGAA 480
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1147 GAAGAAATGAACTCCAAAGACATGACAGATGAAATTTCTTTGGATTTAGGACATCATGAA 1206
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 481 AGGTCTATCATGACAGATCTATACCTCAGTCAAAACAGATGGAGCGGGGATTTGGCGT 540
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1207 AGGTCTATAATGACGAGTCTATACCTCAGTCAAAACAGATGGAGCGGGGATTTGGCGG 1266
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 541 GAARAAGGCCAAGATCTGACAGAGCTGGTCCAGCGGAGAAATTAACATATCTCCAGAAAT 600
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1267 GAAAAGAGGCCAAGATCTGACAGAGCTGGTTCAGCGGAGAAATTAACATATCTCCAGAAAT 1326
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 601 CCTAAGGAGTGCAGCAAAAGCCAGGAGCTGGTGTGTAAACATCAATAAAGGCTGTGGCTAT 660
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1327 CCCAAGGAGTGCAGCAAAAGCCAAGAGCTGGTGTGTAAATCAACAAGGCTGTGGCTAT 1386
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 661 GGTTCCTCAATCCATCACTGCTGTCTACTGTTTTCATGATTTGTCAGTCCAGCGAACA 720
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1387 GGCCTCAGCTCCATCATGCTGTCTACTGCTTTCATGATTTGATATGCGACCCAGCGAACA 1446
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 721 CTCATCTTTGGATCTCAGAAATGGCGCTATGCTACTGTTGGATGGAGCTGTGTTTGA 780
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1447 CTCATCTTTGGATCTCAGAAATTTGGCGCTATGCTACTGTTGGATGGAGCTGTATTTAGG 1506
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 781 CCTGTAAGTGAACATGTACAGACAGATCTGGCCCTCTCCACTGGACACTGGTCAGGTGAA 840
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1507 CCTGTAAGTGAACATGTACAGACAGATCTGGCATCTCCACTGGACACTGGTCAGGTGAA 1566
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 841 GTAATGACAAAAACATTCAGTGTGAGCTCCCAATTTGTAGACAGCTCCATCCTCGG 900
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1567 GTGAAGGACAAAAATTTCAAGTGTGAGCTTCCCAATTTGTAGACAGCTTTCATCCCCGT 1626
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 901 CCTCCTTACTTACCTGCTGTTTCCAGAGACCTTGCAGACCCGACTCTTAAGAGTCCAT 960
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1627 CCTCCATATTTTACCTTGGCTGTACAGAGACCTTCGAGATCGACTTTGTACGAGTGCAT 1686
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 961 GGTGACCCCTGCACTGTTGGTGGTGTCCCAAGTTTGTCAAATCTTGTGATTCGTCCACAACCT 1020
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1687 GGTGACCCCTGCACTGTGGTGGTGTCTCAGTTTGTCAAATCTTTGATCCGCCACAGCT 1746
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1021 TGGCTTGGAAAAAGGAAATAGAAAGAACCCACCAAGAGCTTTGGCTTCAAAACATCCAGTTAT 1080
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1747 TGGCTAGAAAAGAAATAGAAAGAACCCACCAAGAGCTTTGGCTTCAAAACATCCAGTTAT 1806
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1081 GGAGTCCATGTTCAGACGACACAGAAAGTGGGAAACAGAGAGCTTCCACCCCATCGAG 1140
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1807 GGAGTCCATGTTCAGACGACACAGAAAGTGGGAAACAGAGAGCTTCCATCCCATCGAA 1866
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1141 GAGTACATGTTCACGCTTGAAGAAACATTTTCAGCTTCTCGCACGAGAGATGCAAGTGAAT 1200
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1867 GAGTACATGTTCAGCTTGAAGAAACATTTTCAGCTTCTTCGACGAGAGATGCAAGTGGAC 1926
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1201 AAAAAAGAGTATATCTGGCTACTGATGATCCTACTTTTAAAGAGGAGGCAAGACAAAG 1260
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1927 AAAAAAGAGTATTTGGCCACAGATGACCCCTTCTTTTAAAGAGGAGGCAAGACAAAG 1986
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1261 TACTCCAATATGAATTTATTTAGTAACTCTATTTCTTGGTCACTGAGCTGAGTACACAAT 1320
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1987 TACCCCAATATGAATTTATTTAGTAACTCTATTTCTTGGTCACTGAGCTGAGTACACAAT 2046
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1321 CGGTACACAGAAAAATTCATTTGGGGTGTGATCTTGGATATACATTTCTCTCACAGCT 1380
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2047 CGATACACAGAAAAATTCATTTCTGGAGTGTATCTCTGATATACATTTTCTCTCAGGCA 2106
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1381 GACTTCTAGTGTCTACTTTTTCATCCAGCTCTGTCGGGTGCTTATGAATCATGCAA 1440
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2107 GACTTCTAGTGTCTACTTTTTCATCCAGCTCTGTCGGGTGCTTATGAATCATGCAA 2166
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1441 ACCCTGATCTGATGCTCTCGCAACTTCCATTTCTTTGGATGACATCTACTATTTTGA 1500
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2167 ACATCATCTCTGATGCTCTGCAACTTCCATTTCTTTAGATGACATCTACTATTTTGG 2226
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1501 GGCCAAATGCCACATCAGATTGCTTTATCTCAAAACCTCGAACTGAAAGGAA 1560
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2227 GGCCAGAAATGCCACATCAATCAATTTGCTTATGCTCACCACCCGAACTGCAGATGAA 2286
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1561 ATTCCAATGGAACTGAGATATCATTTGCTGCTGGAACCATTTGGGATGGTTATTTCT 1620
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2287 ATTCCTGGAACCTGAGATATCATTTGCTGCTGGAACCATTTGGGATGGTTATTTCT 2346
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1621 AAGGTATCAACAGAAACTTTGGAAAAACAGGCTTATATCCCTCTCAAAAGTCCGAGAG 1680
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2347 AAGGTCTCAACAGAAATTTGGAGGAGCGGCTTATATCCCTCTCAAAAGTCCGAGAG 2406
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1681 AAGTATGAAAACAGTCAAGTATCCACATATCTCTGAAAGCTGAAAAATA 1727
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2407 AAGATAGAAACGGTCAAGTACCCCAATATCTCTGAGGCTGAGAAATA 2453
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
ABV28283
ID ABV28283 standard; cDNA; 3291 BP.
XX
AC ABV28283;
XX
DT 16-SEP-2002 (first entry)
XX
Human prostate expression marker cDNA 28274.
XX
Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
pharmacogenomic marker; gene; ss.
XX
Homo sapiens.
XX
WO200160860-A2.
XX
23-AUG-2001.
XX
20-FEB-2001; 2001WO-US05171.
XX
```

PR 17-FEB-2000; 2000US-183319P.  
PR 16-MAR-2000; 2000US-189862P.  
PR 25-MAY-2000; 2000US-207454P.  
PR 09-JUN-2000; 2000US-211314P.  
PR 18-JUL-2000; 2000US-219007P.  
PR 13-DEC-2000; 2000US-255281P.  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
PA Schlegel R, Endege WO, Monahan JB;  
XX WPI; 2001-662795/76.  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer -  
XX Claim 1; Page 5883; 11750pp; English.  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV0010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX Sequence 3291 BP; 925 A; 729 C; 745 G; 871 T; 21 other;  
SQ Query Match 86.9%; Score 1501.4; DB 23; Length 3291;  
Best Local Similarity 91.8%; Pred. No. 0;  
Matches 1586; Conservative 0; Mismatches 141; Indels 0; Gaps 0;  
QY 1 ATGCGGCGATGACCTGGTTCCTGGGCTGGGATATGCTCATTTCTTTTGGCTGGGGGACC 60  
DB 727 ATGCGGCGATGACCTGGTTCCTGGGCTGGGATATGCTCATTTCTTTTGGCTGGGGGACC 786  
QY 61 TTGTTATTTATATAGGTGGTCAATTTGTTTCGAGATAATGACCCCTGATCCTCCAGC 120  
DB 787 TTGCTGTTTTATATAGGTGGTCAATTTGTTTCGAGATAATGACCCCTGATCCTCTAGC 846  
QY 121 AGAGAACTCTCCAAGATCTTCCAAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 180  
DB 847 CGAGAACTCTCAAGATCTTCCAAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 906  
QY 181 AGGCGAAATGGTGTCTCTCCGAATACAGAAAGCCCTTGAACAGGAGACACTACA 240  
DB 907 AGGCGAAATGGGCGAATCTCTCCGATACCAAGGCGCTTATGATCAGGGGCGAGCTATA 966  
QY 241 GGAAGAGTCCGTTTATAGAAAGACAGCTTGTATAGGCGCAAGACAGATGAAATTAAC 300  
DB 967 GGAAGAGTACGGTTTATAGAAAGACAGCTTGTATAGGCGCAAGACAGATGAAATTAAC 1026  
QY 301 AAGAAACAGCTAGAAATGGTCTGGGAGAGATCATGAAATCTTAAAGAGAGGATTGAA 360  
DB 1027 AAGAAACAGACAGAAATGGTCTGGGAGAGATCATGAAATCTTAAAGAGAGGATTGAA 1086  
QY 361 AATGAGCTTAAAGAGCTCTGGTTTTTCTTCAAGCGAACTGAAGAAATTAAGCAATTA 420  
DB 1087 AATGAGCTTAAAGAGCTCTGGTTTTTCTTCAAGCGAACTGAAGAAATTAAGCAATTA 1146  
QY 421 GAAGGAAATGAATCTCAAGACATCGAGATGAAATTTCTTTGGATTTAGGACACCATGAA 480  
DB 1147 GAAGGAAATGAATCTCAAGACATCGAGATGAAATTTCTTTGGATTTAGGACATCATGAA 1206

QY 481 AGGTCTATCATGACAGATCTATACCTCAGTCAAAACAGATGAGCAGGCGATTGGCGT 540  
DB 1207 AGGTCTATTAATGACGAGTATCTACTCAGTCAAGACAGATGAGCAGGCGATTGGCGG 1266  
QY 541 GAAAAGAGGCGCAAGATCTGACAGAGCTGGTCAGGCGGAGAAATAACATATCTCCAGAA 600  
DB 1267 GAAAAGAGGCGCAAGATCTGACAGAACTGGTCAGGCGGAGAAATAACATATCTCCAGAA 1326  
QY 601 CCTAAGGACTGACAGAAAGCCAGGAAGCTGGTGTAAACATCAATAAAGGCTGGGCTAT 660  
DB 1327 CCCAAGGACTGACAGCAAAAGCCAAAAGCTGGTGTAAATATCAACAAAGGCTGGGCTAT 1386  
QY 661 GGTGTCAACTCCATCAGCTGGTCTACTGTTTCATGATTGCTTATGGCACCAGGCAACA 720  
DB 1387 GGCTGTGAGCTCCATCATGCTGCTACTGCTTCATGATTGCTATGGCACCAGGCAACA 1446  
QY 721 CTCAATCTTGAATCTCAGAAATTTGGCGTATGCTACTGCTGATGGGAGACTGTGTTTGA 780  
DB 1447 CTCAATCTTGAATCTCAGAAATTTGGCGTATGCTACTGCTGATGGGAGACTGTATTAGG 1506  
QY 781 CCTGTAAGTGAGACATGTACAGACAGATCTGGCCTCTCCACTGGACACTGTCAGGTGAA 840  
DB 1507 CCTGTAAGTGAGACATGTACAGACAGATCTGGCCTCTCCACTGGACACTGTCAGGTGAA 1566  
QY 841 GTAATGACAAAAACATTTCAAGTGGTGGTCTCCCATTTGTAGACAGCTCCATCCTCGG 900  
DB 1567 GTGAAGGACAAAAATGTTCAAGTGGTGGTCTCCCATTTGTAGACAGCTCTTCATCCCCG 1626  
QY 901 CCTCTTACTTTACCACCTGGCTGTTCCAGAACACTTGCAGACCGACTCCTAAGAGTCCAT 960  
DB 1627 CCTCCATATTTACCCTTGGCTGTACCAAGAACCTTGCAGATGCACTGTAGAGTGCAT 1686  
QY 961 GGTGACCTGTCAGTGTGGTGGTCTCCCATTTGTCAAATACTTGAATTCGTCACAACT 1020  
DB 1687 GGTGACCTGTCAGTGTGGTGGTCTCTCAGTTTGTCAAATACTTGAATTCGTCACAACT 1746  
QY 1021 TGGCTGGAAGAGAAATAGAAAGACCCACCAAGAGCTTGGCTTCAAACATCCAGTTAT 1080  
DB 1747 TGGCTAGAAAAAGAAATAGAAAGACCCACCAAGAGCTTGGCTTCAAACATCCAGTTAT 1806  
QY 1081 GGAGTCCATGTACAGCCGACAGCAAAAGTGGGAAACAGAGCAGCTTCCACCCCATCGAG 1140  
DB 1807 GGAGTCCATGTACAGCCGACAGCAAAAGTGGGAAACAGAGCTGCTTCCATCCCATGAA 1866  
QY 1141 GAGTACATGTGTACAGTTGAAGAACTTTTCAGCTTCTCGCACGACAGAAATCAAGTGGAT 1200  
DB 1867 GAGTACATGTGTGTGAAGAACTTTTCAGCTTCTTCAGCTTCTTCGACGAGAAATGCAAGTGGAC 1926  
QY 1201 AAAAAAGAGTATATCTGGCTACTGATGATCCTACTTGTTTAAAGGAGGCAAGACAAG 1260  
DB 1927 AAAAAAGAGTATATTTGGCCACAGATGACCTTTCTTTATTAAGGAGGCAAAACAAG 1986  
QY 1261 TACTCCAAATTAAGATTTATTTAGTGATAACTCTATTTCTTGGTTCAGCTGGACTACACAAT 1320  
DB 1987 TACCCCAATTAAGATTTATTTAGTGATACTCTATTTCTTGGTTCAGCTGGACTGCAAT 2046  
QY 1321 CGGTACACAGAAAAATTCACCTTCGGGGTGTGATCTCTGGATATACACTTTCTCTCAGGCT 1380  
DB 2047 CGATACACAGAAAAATTCACCTTCGGGGTGTGATCTCTGGATATACACTTTCTCTCAGGCA 2106  
QY 1381 GACTTCTAGTGTGATCTTTTTCATCCAGGTCTGTGGGTGCTTATGAATCATGCA 1440  
DB 2107 GACTTCTAGTGTGATCTTTTTCATCCAGGTCTGTGGGTGCTTATGAATCATGCA 2166  
QY 1441 ACCCTGCTATCTGATGCTCTCGCAACTTCCATTTCTTTGGATGACATCTACTATTTCGA 1500  
DB 2167 ACCTACATCTGATGCTCTCGCAACTTCCATTTCTTTGGATGACATCTACTATTTCGA 2226  
QY 1501 GGCCAAATGCCCAATTCAGATTTGCTTTTATCTCTCAAAACCTCGAACTGAGAGGAA 1560  
DB 2227 GGCCAGAAATGCCCAATTCAGATTTGCTTTTATGCTCACCACCAACCCGAACTGCAATGAA 2286  
QY 1561 ATTCCAATGGAACTGAGGATATCATTTGGTGGTGGCTGGAACCATTTGGATGGTATTCT 1620

Db 2287 ATTCCCATGGAACCTGGAGATATCATTTGGTGGCTGGAATCATTTGGATGGCTATTCT 2346  
Qy 1621 AAAGGTATCAACAGAAACTTGGAAACAGGCTTATATCCCTCTCAAAAGTCCGAGAG 1680  
Db 2347 AAAGGTGTCAACAGAAATTTGGGAAGGACGGGCTTATATCCCTCTCAAAAGTTCGAGAG 2406  
Qy 1681 AAGATAGAAAACAGTCAAGTATCCACATATATCTCTGAAGCTGAAAAATA 1727  
Db 2407 AAGATAGAAACGGTCNAGTACCCACATATCTCTGAGCTGAGAAATA 2453

## RESULT 7

AAH33315

ID AAH33315 standard; cDNA; 3007 BP.

XX

AC AAH33315;

XX

DT 03-SEP-2001 (first entry)

XX

DE Human colon cancer antigen encoding cDNA SEQ ID NO:371.

XX

KW Human; colon cancer; colon cancer antigen; diagnosis; detection;

KW colorectal carcinoma; chromosome 14; ss.

XX

OS Homo sapiens.

XX

PN WO200122920-A2.

XX

PD 05-APR-2001.

XX

PF 28-SEP-2000; 2000WO-US26524.

XX

PR 29-SEP-1999; 99US-0157137.

XX

PR 03-NOV-1999; 99US-0163280.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Ruben SM, Barash SC, Birse CE, Rosen CA;

XX

DR WPI; 2001-235357/24.

XX

DR P-PSDB; AAG73884.

XX

PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,

XX

PT useful for preventing, diagnosing and/or treating colorectal cancers -

XX

PS Claim 1; Page 2480-2481; 9803pp; English.

XX

CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon

XX

CC cancer-associated nucleic acid molecules (N) and proteins (P), where

XX

CC the proteins are collectively known as colon cancer antigens. The colon

XX

CC cancer antigens have cytostatic activity and can be used in gene

XX

CC therapy and vaccine production. N and P may be used in the prevention,

XX

CC diagnosis and treatment of diseases associated with inappropriate p

XX

CC expression. For example, N and P may be used to treat disorders

XX

CC associated with decreased expression by rectifying mutations or deletions

XX

CC in a patient's genome that affect the activity of p by expressing

XX

CC inactive proteins or to supplement the patients own production of p.

XX

CC Additionally, N may be used to produce the colon cancer-associated Ps,

XX

CC by inserting the nucleic acids into a host cell and culturing the cell

XX

CC to express the proteins. N and P can be used in the prevention, diagnosis

XX

CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204

XX

CC and AAH7789 represent sequences used in the exemplification of the

XX

CC present invention.

XX

CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were

XX

CC missing at time of publication, meaning no sequences are present for

XX

CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX

XX Sequence 3007 BP; 900 A; 597 C; 668 G; 834 T; 8 other;

SQ

Query Match

Best Local Similarity 86.2%; Score 1490; DB 22; Length 3007;

Matches 1585; Conservative 1; Mismatches 141; Indels 1; Gaps 1;

1;

```
QY 1080 TGAAGTCCATGTCAGACGCACAGACAAAGTGGGAACAGAGCAGCTTCCACCCCATCGA 1139
Db |||
QY 1500 TGAAGTCCATGTCAGACGCACAGACAAAGTGGGAACAGAGCAGCTTCCATCCCATGGA 1559
Db |||
QY 1140 GGAGTACATGTTACAGCTTGAAGACATTTTCAGCTTCTCGACGCAGAAATGCCAAGTGA 1199
Db |||
QY 1560 AGAGTACATGTTGTCATGTTGAAGACATTTTCAGCTTCTCGACGCAGAAATGCCAAGTGA 1619
QY 1200 TAAAAAAGAGTATATCTGGCTACTGATGATCCTACTTTGTTTAAAGGAGGCAAGACAAA 1259
Db |||
QY 1620 CAAAAAAGAGTATATTTGGCCACAGATGACCTTCTTTTAAAGGAGGCAAGACAAA 1679
QY 1260 GTACTCCAAATTAAGAAATTAATAGTGAATCACTTATTTCTTGTCAGCTGACGTACAAA 1319
Db |||
QY 1680 GTACCCCAATTAAGAAATTAATAGTGAATCACTTATTTCTTGTCAGCTGACGTACAAA 1739
QY 1320 TCGGTACACAGAAATTCACCTCGGGGTGTGATCTCGATATACACTTCTCTACAGGC 1379
Db |||
QY 1740 TCGATACACAGAAATTCACCTCGGGGTGTGATCTCGATATACACTTCTCTACAGGC 1799
QY 1380 TGACTTTCTAGTGTGCTACTTTTTCATCCAGGTCTGTGGGTGCTTATGAAATCATGCA 1439
Db |||
QY 1800 AGACTTCTAGTGTGCTACTTTTTCATCCAGGTCTGTGGGTGCTTATGAAATTAATGCA 1859
QY 1440 AACCTGTGATCTGATGCTCTGCAACTTCCATTTCTTTGGATGACATCTACTATTTTGG 1499
Db |||
QY 1860 AACACTACATCTGATGCTCTGCAACTTCCATTTCTTTAGATGACATCTACTATTTTGG 1919
QY 1500 AGGCCAAATGCCCAATCAGATGCTGTTTATCTCTCAAACTCCAACTGCAAGTGAAGGA 1559
Db |||
QY 1920 GGGCAGAAATGCCCAATCAGATGCTGTTTATCTCTCAAACTCCAACTGCAAGTGA 1979
QY 1560 AATCCCAATGGAACCTGGAGATATCATTTGGTGTGCTGCAAACTTGGGATGGTTATTC 1619
Db |||
QY 1980 AATCCCAATGGAACCTGGAGATATCATTTGGTGTGCTGCAAACTTGGGATGGTATTC 2039
QY 1620 TAAAGGTATCAACAGAAACTTGGAAACACAGGCTTATATCCCTCTCAAAAGTCCGAGA 1679
Db |||
QY 2040 TAAAGGTGTCAACAGAAACTTGGAAAGGACGGGCTATATCCCTCTCAAAAGTCCGAGA 2099
QY 1680 GNAGATAGAAACAGTCAAGTATCCACATATCTCAAGCTGAAATA 1727
Db |||
QY 2100 GAAAGATAGAAACGGTCAAGTATCCACATATCTCTGAGGCTGAGAAATA 2147
```

## RESULT 8

ID AAT76573 standard; cDNA to mRNA; 1728 BP.

AC AAT76573;

DT 05-MAR-1998 (first entry)

DE Pig alpha 1-6 fucosyltransferase gene.

XX Alpha 1-6 fucosyltransferase; enzyme; pig; human; fucose transfer;  
KW guanosine diphosphate; sugar chain synthesis; modification; antibody;  
KW GlcNAc; cancer diagnosis; SS.

XX Sus scrofa.

XX Key Location/Qualifiers  
FH 1..1728  
FT CDS /tag= a  
FT FT

XX WO9727303-A1.

XX 31-JUL-1997.

XX 23-JAN-1997; 97MO-JP00171.

XX 22-JUL-1996; 96JP-0192260.

XX 24-JAN-1996; 96JP-0010365.

PR 21-JUN-1996; 96JP-0161648.  
PR 24-JUN-1996; 96JP-0162813.  
XX (TOYM ) TOYO BOSEKI KK.

XX Shiba T, Taniguchi N, Uozumi N, Yanagidani S;  
PI WPI; 1997-393690/36.  
XX P-PSDB; AAW22124.

DR Human or pig alpha 1-6 fucosyltransferase and DNA encoding it - for  
PT synthesis and modification of sugar chains and used as an antigen  
PT for production of diagnostic antibodies

XX Claim 5; Page 30-34; 61pp; Japanese.

CC AAT76573 and AAT76574 represent the coding sequences for the pig and  
CC human alpha 1-6 fucosyltransferases of the invention, respectively. The  
CC encoded enzyme transfers fucose from guanosine diphosphate to the  
CC 6-hydroxyl group of the GlcNAc nearest to R in the receptor molecule:  
CC (GlcNAc)beta 1-2Manalalpha 1-6) (GlcNAc)beta 1-2Manalalpha 1-3)Manbeta  
CC 1-4GlcNAc)beta 1-4GlcNAc-R to give (GlcNAc)beta 1-4(Fucalalpha 1-6)GlcNAc-R. It has  
CC an optimum pH of about 7.0 (pig) or 7.5 (human), and is stable over the  
CC pH range 4-10 after 5 hours at 4 degrees C. The optimum working  
CC temperature of the enzyme is 30-37 degrees C. A bivalent metal is not  
CC required for activity of the enzyme, and the enzyme is not inhibited in  
CC the presence of 5 mM EDTA. The enzyme is useful in the synthesis and  
CC modification of sugar chains, and as antigen for the production of  
CC antibodies recognising the enzyme. The antibodies can be used for the  
CC diagnosis of cancer and other diseases.

XX SQ Sequence 1728 BP; 521 A; 362 C; 419 G; 426 T; 0 other;

Query Match 85.0%; Score 1469.4; DB 18; Length 1728;

Best Local Similarity 90.7%; Pred. No. 0;

Matches 1566; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

OY 1 ATGCGGCATCGGACTGGTTCCTGCGGTTGGATTATGCTCATTTCTTTTGGCTGGGGACC 60

Db 1 ATGCGGCATCGGACTGGTTCCTGCGGTTGGATTATGCTCATTTCTTTTGGCTGGGGACC 60

OY 61 TTGTTATTTTATATAGTGGTTCATTTGGTTCGAGATATGACCACTGATCCTCCAGC 120

Db 61 TTGCTATTTTACATAGTGGTTCATTTGGTTCGAGATATGACCACTGATCCTCCAGC 120

OY 121 AGCAACTCTCCAGATTTCTTGCAAGCTTGAAACGCTTAAACACAGCAAAATGAAGACTTG 180

Db 121 CGAAGACTGTCCAAGATTTTGGCAAGCTGGAACGCTTAAACACAAATGAAGACTTG 180

OY 181 AGCGCAATGGCTGAGTCTCTCCGAATACCAAGAGGCCCATTTGACCGGGGACAGCTACA 240

Db 181 AGGAGATGGCTGAATCTCTCCGAATACCAAGAGGCCCATTTGATCAGGGGCCAGCTTCA 240

OY 241 GGAAGAGTCCGTGTTTGTAGAGAAACAGCTTGTAAAGCCAAAGAACAGATGAAAATATAC 300

Db 241 GGAAGAGTCCGTGTTTGTAGAGAAATTTATGAAGGCCAAAGAACAGATGAAAATATAT 300

OY 301 AAGAAACAGCTAGAAATGGTCTCGGGGAAGCATCATGAAATCTTAAAGAGGAGGATTCGA 360

Db 301 AAGAAACAAACTAAAAATGGTCCAGGGAAGGATCATGAAATCTTAAAGAGGAGGATTCGA 360

OY 361 AATGGAGCTAAAGAGCTCTGGTTTTTTCTCAAAAGCGAACTGAAGAAATTTAAAGCATTTA 420

Db 361 AATGGAGCTAAAGAGCTCTGGTTTTTTCTCAAAAGTGAAGTGTGAAGAAATTTAAAGATTTA 420

OY 421 GAAGGAAATGAATCCAAAGACATGCAGATGAATTTCTTTTGGATTTTAGGACACCATGAA 480

Db 421 GAAGGAAATGAATCCAAAGACATGCAGATGAATTTCTTTTATCAGATTTTGGGACATCATGAA 480

OY 481 AGGTCTATCATGACAGATCTATCTACTCTAGTCAAAACAGATGGAGCGGGATTTGGCGT 540

Db 481 AGGTCTATTAATGACGGATCTATCTACTCTAGTCAAAACAGATGGGCGGCGTGTGGCGT 540

Qy	541	GA	AAAAAGAGGCCAAAGATCTGACAGAGCTGGTCCAGCGGAGATAACATATCTCCAGAAT	600
Db	541	GA	AAAAAGAGGCCAAAGATCTGACAGAGCTGGTCCAGCGGAGATAACATATCTCCAGAAT	600
Qy	601	CC	TAAAGACATGCACGAAGCCAGGAAGCTGGTGTGTAACTCAATAAAGGCTGTGGCTAT	660
Db	601	CC	CAAGGACATGCACGAAGCCAGGAAGCTAGTGTGTAACTCAACAAGGCTGTGGCTAT	660
Qy	661	GG	TGTGTCAACTCCATCACTGAGTGTCTACTGTTTTCATGATTCGCTATATGCGCAACCA	720
Db	661	GG	CTGTGTCAGCTCCATCATGTAGTGTACTGCTTTATGATTCGATATGSCACCCAGCGAACA	720
Qy	721	CT	CATCTTTGGAAATCTCAGAAATTTGGCGCTATGCTACTCTGGTGGATGGGAGACTGTGTTT	780
Db	721	CT	CGCCTTTGGAAATCTCAAAATTTGGCGCTACGCTACTGGGGGATGGGAAACTGTGTTT	780
Qy	781	CC	TGTAAGTGTGAGACATGTACAGACAGATCTGGGCTCTCCACTGGACACTGGTCAGGTGAA	840
Db	781	CC	TGTAAGTGTGAGACCTGCACAGACAGATCTGGCAGCTCCACTGGACACTGGTCAGGTGAA	840
Qy	841	GT	AAATGACAAAAACATTTCAAGTGGTCCGAGCTCCCCATTTGTAGACAGCTCCATCCTCGG	900
Db	841	GT	AAAGACAAAAAATGTTTCAGGTGGTTGAGCTCCCAATTTGTAGACAGTGTTCATCCTCGT	900
Qy	901	CT	CTCTTTACTTACACTGGCTGTTCAGAGAAGACCTTCGACAGCGACTCTTAAGAGTCCAT	960
Db	901	CT	CTCCATATTTTACCCCTGGCTGTCCAGAGAAGCCTTCAGATCGACTTGTACGAGTCCAT	960
Qy	961	GG	TGACCTCGAGTGTGGTGTCCAGTGTGTCAAAATACTTTGATTCGTGCACAACTT	1020
Db	961	GG	TGATCCTGAGTGTGGTGTATCCAGTGTTCAGTACTTTGATTCGCCACAACTT	1020
Qy	1021	TG	CTGGAAAGGAAATAGAAAGAGCCACCAAGAGCTTGGCTTCAAACTCCAGTTATT	1080
Db	1021	TG	CTGGAAAGGAAATAGAAAGAGCGCCACCAAGAGCTAGGCTTCAAACTCCAGTTATT	1080
Qy	1081	GG	AGTCCATGTACAGCCGACAGAAAGTGGGAACAGAAAGAGGCTTCCACCCCATCGAG	1140
Db	1081	GG	AGTCCATGTTTAGACGACAGAAAGTGGGAGCGGAAGAGGCTTCCATCCCATTGAG	1140
Qy	1141	GAG	TACATGGTACACGTTGAAGAACTTTTTCAGCTTCTCGCACGCAAAATGCAAGTGGAT	1200
Db	1141	GA	TACACGGTGCAGTTGAAGAAGACTTTCAGCTTCTTCTCGCAGAAATGCAAGTGGAT	1200
Qy	1201	AAAA	AGAGTATATCTGGCTACTGATGATCCTACTTTGTAAAGGAGGCAAGACAAAG	1260
Db	1201	AAAA	AGGGTGTATTTTGGCCACAGATGACCTCTCTTTGTAAAGGAGGCAAAAAAAG	1260
Qy	1261	TACT	CCAAATATGAATTTATATAGTGATAACTCTATTTTCTTGGTCAGCTGACATACACAAT	1320
Db	1261	TAC	CCCCAGTTATGAATTTATATAGTGATACTCTATCTCTTGGTCAGCTGACATACATAAT	1320
Qy	1321	CGGT	TACAGAAAAATTCACCTTGGGGTGTATCCTCGGATATACACTTTCTCTCACAGGCT	1380
Db	1321	CGA	TATACAGAAAAATTCACCTGGGGTGTATCCTCGGATATACACTTTCTCTCCAGGCA	1380
Qy	1381	GACT	TTCTAGTGTACTTTTTCATCCAGGCTGTCTGGGTGCTGTATGAAATCATGCAA	1440
Db	1381	GACT	TTCTAGTGTACTTTTTCATCCAGGCTGTGTAGAGTTGCTTATGAAATCATGCAA	1440
Qy	1441	ACCT	TGATCCTGTATGCTCTGCGAACTTCCATTTCTTGGATGACATCTACTATTGGA	1500
Db	1441	CGCT	TGATCCTGTATGCTCTGCGAACTTCCGTTCTTTTGGATGACATCTACTATTGGA	1500
Qy	1501	GGCC	AAAAATGCCCAACTCAGATTCGTGTTTATCCTCACAAACCTCGAACTGAAGAGGAA	1560
Db	1501	GGCC	AAATGCCCAACCAAAATTTGCCATTTATCCTCACCAACCTCGAACTGAAGAGGAA	1560
Qy	1561	ATT	CCAATGGAACTCGAGATATCAATTTGGTGTGCTGGAAACCAATTCGGATGTTATCT	1620
Db	1561	AT	CCCCATGAAACTCGAGATATTTATTTGGTGTGCTGGAAATCACTGGATGGCTATCT	1620

Qy	1621	AAAGGTATCAACAGAAAACCTTGGAAAAACAGGCTTATATCCCTCTCTACAAGTCCGAGAG	1680
Db	1621	AAAGGTGTTTAAACAGAAAACCTTGGAAAGACGGGCTTATATCCCTCTCTACAAGTTCGAGAG	1680
Qy	1681	AAAGTAAACACAGTCAAGTATCCACATATATCCCTGAAGCTGAAAAATA	1727
Db	1681	AAAGTAGAAACAGTCAAGTATCCACATATATCCCTGAGGCTGACAAGTA	1727
RESULT 9			
ID	AAC63892		
XX	AC	AAC63892 standard; cDNA; 979 BP.	
XX	AC	AAC63892;	
DT	09-FEB-2001	(first entry)	
XX	XX		
DE	Rat FUT8 cDNA.		
KW	Immunologically functional molecule; immune system; immunomodulation;		
KW	glycosylation; fucose; N-acetylglucosamine; cancer; circulatory disease;		
KW	viral infection; bacterial infection; allergy; autoimmune disease;		
KW	inflammation; antibody; rat; FUT8; ss.		
OS	Rattus sp.		
XX	XX		
FN	W0200061739-A1.		
XX	XX		
PD	19-OCT-2000.		
XX	XX		
PF	07-APR-2000; 2000WO-JP02260.		
XX	XX		
PR	09-APR-1999; 99JP-0103158.		
XX	XX		
PA	(KYOW ) KYOWA HAKKO KOGYO KK.		
XX	XX		
PI	Hanai N, Nakamura K, Shoji E, Yamasaki M, Uchida K, Shinkawa T;		
PI	Imabeppu S, Kanda Y, Yamane N, Anazawa H;		
XX	XX		
DR	WPI; 2000-665129/64.		
XX	XX		
PT	Control of activity of antibodies and other immunologically functional		
PT	molecules by addition or removal of fucose from sugar chain for		
PT	diagnosis and treatment of cancer, allergy and other diseases -		
XX	XX		
PS	Example 8; page 73-74; 81pp; Japanese.		
XX	XX		
CC	The invention relates to a method for controlling the activity of an		
CC	immunologically functional molecule (e.g., an antibody) where the		
CC	control is effected by the presence or absence of fucose bound to an		
CC	N-acetylglucosamine residue at the reducing end of the sugar chain on		
CC	the immunologically functional molecule. The invention also relates to		
CC	methods for the diagnosis, prevention or treatment of diseases which		
CC	involve the modified immunologically functional molecule, and agents		
CC	which stimulate the activity of an immunologically functional molecule.		
CC	The methods of the invention are used for the diagnosis, treatment and		
CC	prevention of a broad range of diseases including cancer, circulatory		
CC	disease, viral or bacterial infection, allergy, autoimmune disease and		
CC	inflammation. The present sequence represents a rat FUT8 cDNA isolated		
CC	in an exemplification of the invention.		
XX	XX		
SQ	Sequence 979 BP; 283 A; 223 C; 220 G; 253 T; 0 other;		
Query Match	53.3%; Score 931; DB 21; Length 979;		
Best Local Similarity	96.9%; Pred. No. 1.4e-272;		
Matches	949; Conservative 0; Mismatches 30; Indels 0; Gaps 0;		
Qy	720	ACTCATCTTGGAAATTCAGAAATTTGGCGCTATGCTACTGGTGGATGGGAGACTGTGTTTAG	779
Db	1	ACTCATCTTGGAAATTCAGAAATTTGGCGCTATGCTACTGGTGGATGGGAGACTGTGTTTAG	60
Qy	780	ACCTGTAAAGTGAACATGTACAGACAGACTTGGCCTCTCCACTGGACACTGGTGA	839



Db 61 ACTCTAAGTGAGACATGACAGACAGATCTGGCCCTCTCCACTGGACACTGGTCTAGGTGA 120  
Qy 840 AGTAATAAGCAAAAACATTCAGTGGTCTGAGTCTCCCTCCCTTTAGACAGCTCCATCTCTCG 899  
Db 121 AGTGAATGACAAAATATTCAGTGGTGGAGTCTCCCTTTAGACAGCTTCATCTCTCG 180  
Qy 900 GCCTCTTACTTACCACTGGCTGTTCCAGAAAGACCTTGACAGACCTTCCCTAAAGAGTCCA 959  
Db 181 GCCTCTTACTTACCACTGGCTGTTCCAGAAAGACCTTGACAGACCTTCCCTAAAGAGTCCA 240  
Qy 960 TGGTGACCTGACGTGGTGGTCTCCAGTTTCTCAATCTTCAATCTTCAATCTTCAATCTTCA 1019  
Db 241 TGGTGACCTGACGTGGTGGTCTCCAGTTTCTCAATCTTCAATCTTCAATCTTCAATCTTCA 300  
Qy 1020 TTGGCTGGAAGAAATAGAAAGACCTTCCAGAAAGCTTGGCTTCAAACTTCCAGTTTAT 1079  
Db 301 TTGGCTGGAAGAAATAGAAAGACCTTCCAGAAAGCTTGGCTTCAAACTTCCAGTTTAT 360  
Qy 1080 TGGAGTCCATGTGACAGCGACAGACAAAGTGGGAACAGAGCAGCTTCCACCCCATCGA 1139  
Db 361 TGGAGTCCATGTGACAGCGACAGACAAAGTGGGAACAGAGCAGCTTCCACCCCATCGA 420  
Qy 1140 GGAGTACATGTTACAGTTGAGAACATTTTCAGCTTCTCGACCGACAGATGCAAGTGA 1199  
Db 421 AGAGTACATGTTACAGTTGAGAACATTTTCAGCTTCTCGACCGACAGATGCAAGTGA 480  
Qy 1200 TAAAAAAGAGTATCTGGCTACTGTGATGATCTTCTTAAAGAGGAGGACAAACAAA 1259  
Db 481 TAAAAAAGAGTATCTGGCTACTGTGATGATCTTCTTAAAGAGGAGGACAAACAAA 540  
Qy 1260 GTACTCCAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1319  
Db 541 GTACTCCAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 600  
Qy 1320 TGGGTACACAGAAATTCACCTTCGGGGTGTGATCTCTGATATACACTTCTCTCACAGGC 1379  
Db 601 TGGGTACACAGAAATTCACCTTCGGGGTGTGATCTCTGATATACACTTCTCTCACAGGC 660  
Qy 1380 TGACTTCTTCTAGTGTGACTTTTCTATCCAGGTCTGTGGGTGCTTATGAATCATGCA 1439  
Db 661 TGACTTCTTCTAGTGTGACTTTTCTATCCAGGTCTGTGGGTGCTTATGAATCATGCA 720  
Qy 1440 AACCTGCACTCTGATGCTCTGCAACTTCCATCTTCTTGGATGACATCTACTATTTGG 1499  
Db 721 AACCTGCACTCTGATGCTCTGCAACTTCCATCTTCTTGGATGACATCTACTATTTGG 780  
Qy 1500 AGGCCAAATGCCCACATCAGATTGCTGTTTATCTCTCACAACCTCGAATGAGAGGA 1559  
Db 781 AGGCCAAATGCCCACATCAGATTGCTGTTTATCTCTCACAACCTCGAATGAGAGGA 840  
Qy 1560 AATTCCTCAATGGAACCTGGAGATATCATTTGGTGTGGCTGGAACCACTTGGATGGTTATTC 1619  
Db 841 AATTCCTCAATGGAACCTGGAGATATCATTTGGTGTGGCTGGAACCACTTGGATGGTTATTC 900  
Qy 1620 TAAAGGTATCAACAGAAAACCTTGGAAAACAGGCTTATATCCCTCTCAAAAGTCCGAGA 1679  
Db 901 TAAAGGTATCAACAGAAAACCTTGGAAAACAGGCTTATATCCCTCTCAAAAGTCCGAGA 960  
Qy 1680 GAAGATAGAAACAGTCAAG 1698  
Db 961 GAAGATAGAAACGGTCAAG 979

## RESULT 10

ABK70037

ID ABK70037 standard; cDNA; 979 BP.

XX AC

XX ABK70037;

XX 15-JUL-2002 (first entry)

XX DT

XX DE

XX Antibody production method related cDNA #5.

KW Antibody production; cytostatic; immunomodulator; vasotropic; virucide;  
KW antibacterial; antiinflammatory; antiallergic; allergy; inflammation;  
KW autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO;  
XX tumour; circulatory disease; infection; primer; ss.  
OS Rattus norvegicus.  
XX WO200231140-A1.  
XX 18-APR-2002.  
XX 05-OCT-2001; 2001WO-JP08804.  
XX 06-OCT-2000; 2000JP-0308526.  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX Kanda Y, Satoh M, Nakamura K, Uchida K, Shinkawa T, Yamane N;  
XX Hosaka E, Yamano K, Yamasaki M, Hanai N;  
XX WPI; 2002-340182/37.  
XX Cells producing antibody compositions including antibody fragments and  
XX fusion proteins with Fc domain of antibody, useful for prevention or  
XX treatment of cancer, immune diseases, circulatory diseases and  
XX infections  
XX Example 9; Page 13-14; 314pp; Japanese.  
XX This invention relates to novel method for antibody production  
XX comprising a Chinese hamster ovarian tissue-originated (CHO) cell  
XX transferred with a gene encoding an antibody molecule for producing a  
XX composition comprising an antibody molecule with an Fc domain bonded  
XX to the N-glycoside linkage complex sugar chain. The produced antibody  
XX compositions are drugs for prevention or treatment of diseases  
XX accompanying tumour, allergy or inflammation, autoimmune diseases,  
XX circulatory diseases, and viral and bacterial infections. The  
XX antibodies can be stably produced using the method of the invention  
XX with high binding activity and potency thus leading to high safety and  
XX reduced side effects when applied alone or in combination with other  
XX drugs for therapy. The present sequence represents a nucleotide  
XX molecule used in the method of the invention.

SQ Sequence 979 BP; 283 A; 223 C; 220 G; 253 T; 0 other;

Query Match 53.9%; Score 931; DB 24; Length 979;  
Best Local Similarity 96.9%; Pred. No. 1.4e-272;  
Matches 949; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy 720 ACTCATCTTGGAAATCTCAGAAATTCGGCTATGCTACTGGTGGATGGGAGACTGTTTGG 779  
Db 1 ACTCATCTTGGAAATCTCAGAAATTCGGCTATGCTACTGGTGGATGGGAGACTGTTTGG 60  
Qy 780 ACCTGTAAGTGAGACATGTACAGACATCTGGGCTCTCCACTGGACACTGGTCTAGGTGA 839  
Db 61 ACCTGTAAGTGAGACATGTACAGACATCTGGGCTCTCCACTGGACACTGGTCTAGGTGA 120  
Qy 840 AGTAAATGACAAAAAATTCAGAGTGGTGGAGTCCCATTTAGACAGCTCCATCTCTCG 899  
Db 121 AGTAAATGACAAAAAATTCAGAGTGGTGGAGTCCCATTTAGACAGCTCCATCTCTCG 180  
Qy 900 GCCTCTTACTTACCACTGGCTGTTCCAGAAAGACCTTTCAGAGACCTCTTAAAGAGTCCA 959  
Db 181 GCCTCTTACTTACCACTGGCTGTTCCAGAAAGACCTTTCAGAGACCTCTTAAAGAGTCCA 240  
Qy 960 TGGTGACCTGACGTGGTGGTCTCCAGTTTCTCAATCTTCAATCTTCAATCTTCAATCTTCA 1019  
Db 241 TGGTGACCTGACGTGGTGGTCTCCAGTTTCTCAATCTTCAATCTTCAATCTTCAATCTTCA 300  
Qy 1020 TTGGCTGGAAGAAATAGAAAGACCTTCCAGAAAGCTTGGCTTCAAACTTCCAGTTTAT 1079  
Db 301 TTGGCTGGAAGAAATAGAAAGACCTTCCAGAAAGCTTGGCTTCAAACTTCCAGTTTAT 360







```

Db 901 TAAAGGTGTCAACAGAAAACTAGAAAAACAGCGCTGTACCTTCTCAAAAGTCCGAGA 960
QY 1680 GAAGATAGAAACAGTCAAG 1698
Db 961 GAAGATAGAAACGGTCAAG 979

RESULT 13
AAF87952
ID AAF87952 standard; DNA; 1017 BP.
AC AAF87952;
XX
DT 20-JUL-2001 (first entry)
XX
DE Human alpha 1-6 fucosyltransferase 237-575 encoding DNA SEQ ID NO:1.
XX
KW Human; alpha 1-6 fucosyltransferase; alpha 1-6 fucT; antibody;
KW alpha 1-6 fucose transferase; anti-human alpha 1-6 fucT; immunoassay;
KW ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1017
FT FT /*tag= a
FT FT /partial
FT FT /product= "alpha 1-6 fucosyltransferase 237-575"
XX
PN JP2001011097-A.
XX
PD 16-JAN-2001.
XX
PF 29-JUN-1999; 99JP-0183569.
XX
PR 29-JUN-1999; 99JP-0183569.
XX
PA (FJRE ) FUJIREBIO KK.
XX
DR WPI; 2001-275926/29.
DR P-PSDB; AAF875061.
XX
PT Novel anti-human alpha1-6 fucose transferase antibody useful for
PT immunoassay -
XX
PS Example 1; Page 6-7; 11pp; Japanese.
XX
CC The present invention describes an anti-human alpha 1-6
CC fucosyltransferase (alpha 1-6 fucT, also called alpha 1-6 fucose
CC transferase) antibody. Also described are: (1) a hybridoma producing
CC the above monoclonal antibody; (2) an immunoassay for detecting human
CC alpha 1-6 fucT by using the above antibody or its antibody fragment;
CC and (3) a reagent used for the above immunoassay. The anti-human
CC alpha 1-6 fucT antibody can be used for immunoassay. The present
CC sequence encodes the human alpha 1-6 fucT protein of residues 237 to
CC 575, which is used in an example from the present invention.
XX
SQ Sequence 1017 BP; 296 A; 231 C; 228 G; 262 T; 0 other;

Query Match 50.7%; Score 876.2; DB 22; Length 1017;
Best Local Similarity 91.3%; Pred. No. 6.8e-256;
Matches 929; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 709 ACCGAGCGAACAACACTCATCTTGGAAATCTCAGAAATGGCGCTATGCTACTGGTGGATGGAG 768
Db 1 ACCGAGCGAACAACACTCATCTTGGAAATCTCAGAAATGGCGCTATGCTACTGGTGGATGGAG 60
QY 769 ACTGTGTTTAGACCTGTAGTGAAGATGTACAGACAGATCTGGCTCTCCACTGGACAC 828
Db 61 ACTGTATTTAGGCGCTGTAGTGAAGATGTACAGACAGATCTGGCTCTCCACTGGACAC 120
QY 829 TGGTCAGGTGAAGTAAATGACAAAAACATTCAGGTGGTGGCTCCCGCTCCCGCTGACACG 888

```

```

Db 121 TGGTCAGGTGAAGTGAAGGACAAAAATGTTCAAGTGGTGGAGCTTCCCATTTAGACACT 180
QY 889 CTCCATCTCGGCGCTCTTACTTACCACTGGCTGTTCAGAGAGACCTTGCAGACCGACTC 948
Db 181 CTTTCATCCCGCTCTCCATATTTACCCCTGGCTGTACCAGAGAGACCTCGCAGATGCATT 240
QY 949 CTAAGAGTCCATGGTGACCCCTGCAGTGTGGTGGTGTCCCAAGTTTGTCAAATPACTTGATT 1008
Db 241 GTACAGTGCATGGTGACCCCTGCAGTGTGGTGGTGTCTCAGTTTGTCAAATPACTTGATC 300
QY 1009 CGTCCAAACCTTGGCTGGAAAAAGAAATAGAAAGCCACCAAGAGCTTGGCTTCAA 1068
Db 301 CGCCCAACAGCTTGGCTAGAAAAAGAAATAGAAAGAGCCACCAAGAGCTTGGCTTCAA 360
QY 1069 CATCCAGTTATTTGGAGTCCATGTCCAGCGCAGACAAAGTGGGAAACAGACACCGCTTC 1128
Db 361 CATCCAGTTATTTGGAGTCCATGTCCAGCGCAGACAAAGTGGGAAACAGAGCTCGCTTC 420
QY 1129 CACCCATCGAGGAGTACATGGTACACGTTGAAGAAACATTTTCAGCTTCTCCGACGCGA 1188
Db 421 CATCCATTTGAAGAGTACATGGTGCATGTTGAAGAACATTTTCAGCTTCTTCGACGCGA 480
QY 1189 ATGCAAGTGGATAAAAAAGAGATATCTGGCTACTGATGATCTCTTTGTTAAAGGAG 1248
Db 481 ATGCAAGTGGACAAAAAAAGAGTGTATTTGGCCACAGATGACCTTCTTTATTAAGGAG 540
QY 1249 GCAAAGACAAAGTACTCCAATTATGAATTTATTAGTGAATACTTATTTCTTGGTCAGCT 1308
Db 541 GCAAAAACAAAGTACCCCAATTTATGAATTTATTAGTGAATACTTATTTCTTGGTCAGCT 600
QY 1309 GGACTACACAATCGGTACACAGAAAAATTCACCTCGGGGTGTGATCCTCGATATACACTTT 1368
Db 601 GGACTGCACAATCGATACACAGAAAAATTCACCTCGGGAGTATCTCGATATACACTTT 660
QY 1369 CTCTCACAGGCTGACTTTCTAGTGTGTAATTTTTCATCCAGGTCTGTGGGTTCCTTAT 1428
Db 661 CTCTCTCAGGCAGACTTCTTAGTGTGTAATTTTTCATCCAGGTCTGTGGGTTCCTTAT 720
QY 1429 GAAATCATGCAACCCCTGCATCCTGATGCTCTCGAACTTCCATCTTTGGATGACATC 1488
Db 721 GAAATTCATGCAACCACTACATCCTGATGCTCTCGAACTTCCATCTTTTAGATGACATC 780
QY 1489 TACTATTTTGGAGCGCAAAATGCCACAATCAGATTCGTGTTTATTCCTCAACAACTCGA 1548
Db 781 TACTATTTTGGAGCGCAAAATGCCACAATCAGATTCGTGTTTATTCCTCAACAACTCGA 840
QY 1549 ACTGAAGAGGAAATTCCAATGGAACCTGGAGATATCATTTGGTGTGGCTGGAAACCATGG 1608
Db 841 ACTGCAGATGAAATTCCTCATGGAACCTGGAGATATCATTTGGTGTGGCTGGAAATCATGG 900
QY 1609 GATGGTTATTTCTAAAGGTATCAACAGAAAACTTGGAAAAACAGGCTTATATCCCTCTTAC 1668
Db 901 GATGGCTATTTCTAAAGGTGTCAACAGGAAATTTGGGAAGGAGCGGCTATATCCCTCTTAC 960
QY 1669 AAGTCCGAGAGAGATAGAACAGTCAAGTATCCACATATCCATATCCTGAGAGTGAATAA 1725
Db 961 AAGTTCGAGAGAGATAGAACCGTCAAGTACCCCAATATCTCTGAGGCTGAGAAA 1017

RESULT 14
AAF87953
ID AAF87953 standard; DNA; 699 BP.
XX
AC AAF87953;
XX
DT 20-JUL-2001 (first entry)
XX
DE Human alpha 1-6 fucosyltransferase 343-575 encoding DNA SEQ ID NO:2.
XX
KW Human; alpha 1-6 fucosyltransferase; alpha 1-6 fucT; antibody;
KW alpha 1-6 fucose transferase; anti-human alpha 1-6 fucT; immunoassay;
KW ds.

```

```
XX OS Homo sapiens.
XX FH Key
XX PH Location/Qualifiers
XX CDS 1..699
XX FT /*tag= a
XX FT /partial
XX FT /product= "alpha 1-6 fucosyltransferase 343-575"
XX PN JP2001011097-A.
XX PD 16-JAN-2001.
XX PF 29-JUN-1999; 99JP-0183569.
XX PR 29-JUN-1999; 99JP-0183569.
XX PA (FURE ) FUJIREBIO KK.
XX DR WPI; 2001-275926/29.
XX DR P-PSDB; AAB75062.
XX PT Novel anti-human alpha1-6 fucose transferase antibody useful for
XX PT immunosassay -
XX PS Example 1; Page 7-8; 11pp; Japanese.
XX CC The present invention describes an anti-human alpha 1-6
XX CC fucosyltransferase (alpha 1-6 fuct, also called alpha 1-6 fucose
XX CC transferase) antibody. Also described are: (1) a hybridoma producing
XX CC the above monoclonal antibody; (2) an immunoassay for detecting human
XX CC alpha 1-6 fuct by using the above antibody or its antibody fragment;
XX CC and (3) a reagent used for the above immunoassay. The anti-human
XX CC alpha 1-6 fuct antibody can be used for immunoassay. The present
XX CC sequence encodes the human alpha 1-6 fuct protein of residues 343 to
XX CC 575, which is used in an example from the present invention.
XX SQ Sequence 699 BP; 220 A; 152 C; 147 G; 180 T; 0 other;

Query Match 35.0%; Score 604.6; DB 22; Length 699;
Best Local Similarity 91.6%; Pred. No. 2.9e-173;
Matches 640; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 1027 GAAAGGAATAGAGAGCCACCAAGAGCTTGGCTTCAACATCCAGTATTGAGTC 1086
DB 1 GAAAGGAATAGAGAGCCACCAAGAGCTTGGCTTCAACATCCAGTATTGAGTC 60
QY 1087 CATGTACAGCCACAGACAAAGTGGAGACAGAGCAGCTTCCACCCATCGAGGATAC 1146
DB 61 CATGTACAGCCACAGACAAAGTGGAGACAGAGCTTCCATCCCATTTGAAGAGTAC 120
QY 1147 ATGGTACAGCTTGAAGAACATTTTCAGCTTCTCGCAGCAGAAATGCAAGTGGATAAAAA 1206
DB 121 ATGGTGCATGTTGAAGAACATTTTCAGCTTCTTGCAGCAGAAATGCAAGTGGACAAAAA 180
QY 1207 AGAGTATATCTGGCTACTGATCCTACTTTGTTTAAAGGAGGCAAAAGCAAAAGTACTCC 1266
DB 181 AGAGTGTATTGGCCACAGATGACCTCTTTTATTAAAGGAGGCAAAAGCAAAAGTACTCC 240
QY 1267 AATTATGAATTTATTAGTGATTAACCTATTCTTGGTCACTGGAGTACACAAATCGGTAC 1326
DB 241 AATTATGAATTTATTAGTGATTAACCTATTCTTGGTCACTGGAGTACACAAATCGGTAC 300
QY 1327 ACAGAAAAATTCATTTGGGGTGTGATCTGTGGATATACACTTCTCTCAGAGGCTCACTTT 1386
DB 301 ACAGAAAAATTCATTTGGGGTGTGATCTGTGGATATACACTTCTCTCAGAGGCTCACTTT 360
QY 1387 CTAGTGTGTACTTTTTCATCCAGGCTGTGTGGGTGCTTATGAATCATGCAAAACCCCTG 1446
DB 361 CTAGTGTGTACTTTTTCATCCAGGCTGTGTGGGTGCTTATGAATCATGCAAAACACTA 420
QY 1447 CATCTGTATGCTCTGCGAGACTTCCATCTTTGGATGACATCTACTATTTTGGAGGCCAA 1506
DB CATCTGTATGCTCTGCGAGACTTCCATCTTTGGATGACATCTACTATTTTGGAGGCCAA 1506

421 CATCTGTATGCTCTGCGAGACTTCCATCTTTAGATGACATCTACTATTTTGGGGGCCAG 480
1507 AATGCCACAAATCAGATTGCTGTTTATCTCCTCACAACCTCGAACTGAAGAGAAATTCCA 1566
481 AATGCCACAAATCAAAATTGCCATTATGCTCACCACACCCCGAACTGCAGATGAATCCC 540
1567 ATGGAACCTGGAGATATCATTTGCTGCTCGAAACCAATTCGGATGGTATTCTTAAAGGT 1626
541 ATGGAACCTGGAGATATCATTTGCTGCTCGAAACCAATTCGGATGGTATTCTTAAAGGT 600
1627 ATCAACAGAAAACTTGGAAAAACAGGCTTATATCCTCTCTCAAAAGTCCGAGAGAAGATA 1686
601 GTCAACAGGAAATTGGGAGGACGCGCTATATCCTCTCTCAAAAGTTCGAGAGAAGATA 660
1687 GAAACAGTCAAGTATCCACATATCTCTGAAGCTGAAAAA 1725
661 GAAACGGTCAAGTATCCACATATCTCTGAGGCTGAGAA 699

RESULT 15
ABL04601
ID ABL04601 standard; cDNA; 2761 BP.
AC ABL04601;
XX
XX DT 26-MAR-2002 (first entry)
XX DE
XX KW Drosophila melanogaster expressed polynucleotide SEQ ID NO 8285.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ss.
XX OS
XX PN Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li FWD, Myers EW;
XX WPI; 2001-656860/75.
XX P-PSDB; ABB60498.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Claim 1; SEQ ID NO 8285; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (AB16176-AB130511), the expressed DNA
XX CC sequences (AB101840-AB16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 2761 BP; 657 A; 731 C; 755 G; 618 T; 0 other;

Query Match 21.9%; Score 378.8; DB 23; Length 2761;
Best Local Similarity 57.2%; Pred. No. 3.9e-104;
Matches 711; Conservative 0; Mismatches 522; Indels 9; Gaps 1;
```

Qy	473	ACCATGAAGGTCATCATCAGACATCTATCTACCTCAGTCTCAAAACAGATGGAGCAGGG	532
Db	1083	AGCACAAGCGCTCCCTCTGAGCGGATATGGAGCGGATGCGTCACTGCGATGGCTACGAGG	1142
Qy	533	ATTGGCGTGAANAAGAGCGCCAAAGATCTGACAGAGCTGCTCCAGCGGAGAATAACATATC	592
Db	1143	CCTGGCGACACAAGAGCGACGGGATTTGACGATTTGGTACAGCGCGCTGCACCATC	1202
Qy	593	TCCAGAAATCCTTAAGGATCTGACGACAAAGCCAGGAAGCTGGTGTGTAAATCAATAAAGGCT	652
Db	1203	TACAGAAATCCAGTGAATTCGCAGAAATGCTCGAAGCTGGTCTGTAAAGCTCAACAAGGCT	1262
Qy	653	GTGGCTATGGTTGTCAACTCCATCACTGGTCTACTGTTTTCATGATTTGCTTATGGCAACC	712
Db	1263	GTGGCTACGGCTGCAACTGCAACCATGTGGTGTATTGCTTTATAGTGGCCCTACGCCACC	1322
Qy	713	AGCGAAACATCATCTCTGGAATCTCAGAAATTTGGGCTATGCTACTGGTGGATGGGAGACTG	772
Db	1323	AGCGAGCGTCACTTTAAAATCCCGCGATGCGGATATCAAGAGGCGGCTGGAGGAGG	1382
Qy	773	TGTTTAGACCTGTAAGTGAGACATGTACAGACAGATCTGGCCTCTCCACTGGAACACTGGT	832
Db	1383	TGTTTCCAACCGGTGTCCAAACAGCTGCCATGATGCGGSCACGGCAACACATACAAATTGGC	1442
Qy	833	CAGTGAAGTAATGACAAAAAATTCAGTGTGAGCTCCCAATTTGTATAGACAGACTCC	892
Db	1443	CGGGCA-----AGCCGAACACCCAGGTGCTGGTGTGGCCCATCAITTGACTCGCTGA	1493
Qy	893	ATCCTCGGCCTCCTTACTTACCACCTGCTGTCCAGAAAGACCTTGCAGACCGACTCCATA	952
Db	1494	TGCCGAGACCGCGGTACTCGCGCTCGCCGTTCCGAGGATCTGGCCCGAGACTTAAAGC	1553
Qy	953	GAGTCCATGGTGACCCCTGCAGTGTGGTGTGCCAGTTTGTCAAAATCTTGAATCTGTC	1012
Db	1554	GTCTGCATGAGACCCCATCGTTTGGTGGGTGGGTCAGTTTCTTAAGTATTTGCTCGCAC	1613
Qy	1013	CACAACTTGGCTGGAAAAGGAATAGAAAGCCACCAAGAGCTTGGCTTCAACATC	1072
Db	1614	CGCAACCAACGACCGGGATTTTCTTACTCTTGCCATGCGCAATTTGGGTGGGAGGCTC	1673
Qy	1073	CAGTTATTGGAGTCCATGTCCAGCGCACAGACAAAGTGGGAAACAGAAAGAGCCCTTCCACC	1132
Db	1674	CCATTGTTGGCTTCATGTCGTGCGACGACAAAGTGGGCACAGAAAGCGGCTGCCACA	1733
Qy	1133	CCATCGAGGAGTACATGGTACACTTGAAGAACATTTTTCAGTCTTCGCAAGCAGAAATGC	1192
Db	1734	GTGTGGAGGAGTATATGACCTATGTCAGAGACTACTACCGCACTGGAGGTAAACGGCA	1793
Qy	1193	AAGTGGATAAAAAAGAGTATATCTGGCTACTGATGATCTACTTTGTTAAAGAGGCA	1252
Db	1794	GCACTGTGGCTCGCCGGAATTTCTCGCATCGGATGATGCCAGGTTATTGAGAGAGCGC	1853
Qy	1253	AGACAAAGTACTCCAAATTATGAATTTATTAGTGATAACTCTATTCTTGTGTCAGCTGGAC	1312
Db	1854	GCCGAAGTACCCGCACTACAGATCATTTGTGATCGGAGGTGGCGGCATGGCGTCCG	1913
Qy	1313	TACAATACGGTACACAGAAAATTCATCTGGGGGTGTGATCTCTGGATATACACTTTCTCT	1372
Db	1914	TATCCACACGATACACGGACACCCGCTGAAACGGGATCATTTCTGGATATCCACTGCTTT	1973
Qy	1373	CACAGGCTGACTTCTAGTGTGTACTTTTTCATCCAGGTCGTGCGGTTGCTTATGAAA	1432
Db	1974	CCATGTCGATCATCTGGTGTGACCTTCTCGTCGAGGTGTGCGGTGGCTTACGAGA	2033
Qy	1433	TCATGCAAAACCTGCATCTGATGCCTCTGCAACTTCCATTTCTTTTGGATGACATCTACT	1492
Db	2034	TAATGCACGATGATCCGGATGACGGCATCGGTTCAAGTCGCTGACGACATATACT	2093
Qy	1493	ATTTTGGAGGCCAAAATGCCCAAAATCAGATTTGCTGTTTATTCCTCAAAAACCTCGAACTG	1552
Db	2094	ACTACGCTGGCAGAAATGCGCAAAATCGCGCGCTCGTTATTCGCCCAACAGCGCGACGC	2153

Search completed: February 2, 2004, 12:33:55  
Job time : 475.615 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 2, 2004, 12:05:46 ; Search time 104.069 Seconds  
(without alignments)  
7328.922 Million cell updates/sec

Title: US-09-971-773-2

Perfect score: 1728

Sequence: 1 atcgggcgatgactgttc.....atcctgaagctgaaaaatag 1728

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1496.6	86.6	2100	3	US-08-913-805A-9
2	1496.6	86.6	2100	3	US-09-442-629-9
3	1469.4	85.0	1728	3	US-08-913-805A-1
4	1469.4	85.0	1728	3	US-09-442-629-1
5	74	4.3	7218	1	US-08-232-463-14
6	38.2	2.2	921	4	US-09-328-352-2691
7	38	2.2	1230025	4	US-09-198-452A-1
8	37.6	2.2	2116	3	US-09-000-041A-1
9	37.2	2.2	1596	4	US-09-134-001C-2083
10	36.8	2.1	1750	2	US-08-933-750C-54
11	36.8	2.1	1750	3	US-09-234-613-54
12	36	2.1	4517	3	US-09-140-804-9
13	36	2.1	4517	4	US-09-686-838B-9
14	36	2.1	4517	4	US-09-776-976-5
15	36	2.1	4517	4	US-09-909-547-5
16	35.8	2.1	1664976	4	US-08-916-421B-1
17	35.2	2.0	978	4	US-09-134-001C-110
18	35	2.0	313	4	US-09-313-294A-4658
19	34.6	2.0	1664976	4	US-08-916-421B-1
20	34.2	2.0	602	4	US-09-538-709-21
21	34.2	2.0	1268	4	US-09-538-709-19
22	34	2.0	1591	2	US-08-646-981-14
23	34	2.0	2915	4	US-09-336-115C-5
24	33.8	2.0	805	3	US-08-961-083-139
25	33.8	2.0	805	4	US-09-536-784-139
26	33.8	2.0	969	4	US-09-252-991A-12688
27	33.8	2.0	1041	4	US-09-252-991A-12831

28	33.8	2.0	1629	4	US-09-252-991A-13111	Sequence 13111, A
29	33.8	2.0	12127	4	US-08-961-527-148	Sequence 148, Appl
30	33.4	1.9	2628	1	US-08-143-219-1	Sequence 1, Appli
31	33.2	1.9	3356	3	US-09-379-523-4	Sequence 4, Appli
32	33	1.9	1383	4	US-09-134-001C-2485	Sequence 2485, Ap
33	32.8	1.9	241	4	US-09-397-787-70	Sequence 70, Appl
34	32.8	1.9	1425	4	US-09-592-998C-6	Sequence 6, Appli
35	32.8	1.9	1590	4	US-08-887-534A-73	Sequence 73, Appl
36	32.8	1.9	1590	4	US-09-527-431-73	Sequence 73, Appl
37	32.8	1.9	1656	3	US-08-699-103B-7	Sequence 7, Appli
38	32.8	1.9	1656	4	US-09-229-059-7	Sequence 7, Appli
39	32.8	1.9	1656	4	US-09-628-133-7	Sequence 7, Appli
40	32.8	1.9	2699	4	US-09-336-115C-3	Sequence 3, Appli
41	32.8	1.9	5361	3	US-08-973-462-2	Sequence 2, Appli
42	32.8	1.9	6152	3	US-08-973-462-1	Sequence 1, Appli
43	32.8	1.9	36651	4	US-09-738-894A-3	Sequence 3, Appli
44	32.8	1.9	36651	4	US-09-964-469-3	Sequence 3, Appli
45	32.8	1.9	168575	4	US-09-426-290-1	Sequence 1, Appli

#### ALIGNMENTS

#### RESULT 1

US-08-913-805A-9  
; Sequence 9, Application US/08913805A  
; Patent No. 6054304  
; GENERAL INFORMATION:  
; APPLICANT: TANIGUCHI, Naoyuki  
; APPLICANT: UOZUMI, Naofumi  
; APPLICANT: SHIBA, Tetsuo  
; APPLICANT: YANAGIDANI, Shusaku  
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3+ Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
; SOFTWARE: WordPerfect 5.1 Windows  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/913.805A  
; FILING DATE: 7 JAN 1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP97/00171  
; FILING DATE: 23 JAN 1997  
; APPLICATION NUMBER: JP 192260  
; FILING DATE: 22 JUL 1996  
; APPLICATION NUMBER: JP 162813  
; FILING DATE: 24 JUN 1996  
; APPLICATION NUMBER: JP 161648  
; FILING DATE: 21 JUN 1996  
; APPLICATION NUMBER: JP 10365  
; FILING DATE: 24 JAN 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toffenetti, Judith L.  
; REGISTRATION NUMBER: 39,048  
; REFERENCE/DOCKET NUMBER: 2356/3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-429-1776  
; TELEFAX: 202-429-0796  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2100 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double



; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-913-805A-9

Query Match 86.6%; Score 1496.6; DB 3; Length 2100;  
Best Local Similarity 91.7%; Pred. No. 0; Mismatches 144; Indels 0; Gaps 0;  
Matches 1583; Conservative 0;

QY 1 ATGCGGCGATGAGTGGTTCCTGGGCTTGGATTAATGCTCATTTCTTTTTCCTGGGGGACC 60  
DB 198 ATGCGGCGATGAGTGGTTCCTGGGCTTGGATTAATGCTCATTTCTTTTTCCTGGGGGACC 257

QY 61 TTGTTATTTTATATAGTGGTCAATTTGGTTCAGATAATGACCACTTCATCTCCAGC 120  
DB 258 TTGCTGTTTTATATAGTGGTCAATTTGGTTCAGATAATGACCACTTCATCTCCAGC 317

QY 121 AGAGAACTCTCCAAAGTCTTCCAAAGCTTGAACGCTTAAACACACAAATGAAGACTTG 180  
DB 318 CGAGAACTCTCCAAAGTCTTCCAAAGCTTGAACGCTTAAACACACAAATGAAGACTTG 377

QY 181 AGGCGAATGGCTGAGTCTCTCCGAATACCAAGAGCCCATTTGACAGGCGGACAGCTACA 240  
DB 378 AGGCGAATGGCGAATCTCTCCGATACCAAGAGCCCATTTGATCAGGCGGCGGACGATA 437

QY 241 GGAAGAGTCCGCTGTTTTAGAGAAACAGCTTGTAAAGGCGGCAAGAAACAGATTGAAAATTAC 300  
DB 438 GGAAGAGTACGCGTTTTAGAGAGAGCTTGTAAAGGCGGCAAGAAACAGATTGAAAATTAC 497

QY 301 AAGAAACAGCTAGAAATGGTCTGGGAGGATCATGAATCTTAAAGAGGAGGATTGAA 360  
DB 498 AAGAAACAGAGCAGAAATGGTCTGGGAGGATCATGAATCTTAAAGAGGAGGATTGAA 557

QY 361 AATGAGCTTAAAGAGCTCTGGTTTTTCTACAAAGCGAACTGAAGAAATTAAGCATTTA 420  
DB 558 AATGAGCTTAAAGAGCTCTGGTTTTTCTACAGAGTGAATTAAGAACTTA 617

QY 421 GAAGAAATGAATCTCAAAGACATGACAGATGAATTTCTTTGGATTTAGGACCATGAA 480  
DB 618 GAAGAAATGAATCTCAAAGACATGACAGATGAATTTCTTTGGATTTAGGACCATGAA 677

QY 481 AGGTCTATCATCAGAGTCTATCTACCTCAGTCAACAGATGGAGCGGAGTGGCGT 540  
DB 678 AGGTCTATCATCAGAGTCTATCTACCTCAGTCAACAGATGGAGCGGAGTGGCGG 737

QY 541 GAAAGAGGCGCAAGAGCTGACAGAGTGGTCCAGCGGAGAAATTAACATATCTCCAGAA 600  
DB 738 GAAAGAGGCGCAAGAGCTGACAGAACTGGTTTACGCGGAGAAATTAACATATCTCCAGAA 797

QY 601 CTAAGGAGTGCAGAAAGCCAGGAGCTGGTGTGTAAACATCAATAAAGGCTGTGGCTAT 660  
DB 798 CCAAGGAGTGCAGAAAGCCAGGAGCTGGTGTGTAAATCAACAAAGGCTGTGGCTAT 857

QY 661 GTTGTCAATCCATCAGTGTCTACTGTTTCATGATTTGTCACCCAGCGGACCA 720  
DB 858 GGTGTCAATCCATCAGTGTCTACTGTTTCATGATTTGTCACCCAGCGGACCA 917

QY 721 CTCATCTTGGAACTCAGAAATTTGGCGCTATGCTACTGTTGGATGGAGAGCTGTGTTAGA 780  
DB 918 CTCATCTTGGAACTCAGAAATTTGGCGCTATGCTACTGTTGGATGGAGAGCTGTGTTAGA 977

QY 781 CTTGTAAGTGTAGACATGTACAGACAGATCTGGCGCTCTCCATCTGGACACTGGTGAG 840  
DB 978 CTTGTAAGTGTAGACATGTACAGACAGATCTGGCGCTCTCCATCTGGACACTGGTGAG 1037

QY 841 GTAAATGACAAACATCTCAAGTGTCCAGCTCCCATTTGTAGACAGCTCCATCTCGG 900  
DB 1038 GTAAATGACAAACATCTCAAGTGTCCAGCTCCCATTTGTAGACAGCTCCATCTCGG 1097

QY 901 CCTCTCTACTTACCACTGGCTCTTCCAGAAAGACCTTGCAGACCGGACTCTTAAGAGTCCAT 960  
DB 1098 CCTCTCTACTTACCACTGGCTCTTCCAGAAAGACCTTGCAGACCGGACTCTTAAGAGTCCAT 1157

QY 961 GGTGACCCCTGACGTGGTGGGTGCCCAAGTTTGTCAAAATCTTGATTCGTGCCCAACCT 1020

DB 1158 GGTGACCCCTGACGTGGTGGTCTCAGTTTGTCAAAATCTTGATCCGCCACAGCT 1217

QY 1021 TGGCTGGAAAAGGAAATAGAGAACCCACCAAGAGCTTGGCTTTCAAACATCCAGTTATT 1080

DB 1218 TGGCTAGAAAAGGAAATAGAGAACCCACCAAGAGCTTGGCTTTCAAACATCCAGTTATT 1277

QY 1081 GGAGTCCATGTACAGCGCACAGACAAAGTGGGAAACAGAGAGGCTTCCACCCCATCGAG 1140

DB 1278 GGAGTCCATGTACAGCGCACAGACAAAGTGGGAAACAGAGAGGCTTCCATCCCATGAA 1337

QY 1141 GAGTACATGGTACAGCTTGAAGAACATTTTCAGCTTCTCGCACGAGAACTCAAGTGCAT 1200

DB 1338 GAGTACATGGTGCATGTGAAGAACATTTTCAGCTTCTTCGCCGAGAACTCAAGTGCAT 1397

QY 1201 AAAAAAGAGATATATCTGGCTACTGATGATCTTACTTTTAAAGAGGCGGCAAGAAAG 1260

DB 1398 AAAAAAGAGTATTTGGCCACAGATGACCTTCTTTTAAAGAGGCGGCAAGAAAG 1457

QY 1261 TACTCCAATTATGAATTTATAGTGAATCTCTATTTCTTGGTCAAGTGGAGCTACAAAT 1320

DB 1458 TACCCCAATTATGAATTTATAGTGAATCTCTATTTCTTGGTCAAGTGGAGCTGACAAAT 1517

QY 1321 CGGTACACAGAAATTCACCTTGGGCTGATCTGGATATACACTTTCTCTCAGAGCT 1380

DB 1518 CGATACACAGAAATTCACCTTGGGAGTGAATCTGGATATACACTTTCTCTCAGGCA 1577

QY 1381 GACTTTCTAGTGTACTTTTCAATCCAGGCTCTGTGGGTTGCTTATGAATCATGCAA 1440

DB 1578 GACTTTCTAGTGTACTTTTCAATCCAGGCTCTGTGGGTTGCTTATGAATCATGCAA 1637

QY 1441 ACCGTGATCTGATGCTCTCGGAACTTCCATTTCTTTGGATGACATCTACTATTTTGG 1500

DB 1638 ACCTGATCTGATGCTCTCGGAACTTCCATTTCTTTAGATGACATCTACTATTTTGG 1697

QY 1501 GGCCAAAATGCCCAATCAGATTGCTTTATCTCACAACCTCGAACTGAAAGAGAA 1560

DB 1698 GGCCAAAATGCCCAATCAGATTGCTTTATCTCACAACCTCGAACTGAAAGAGAA 1757

QY 1561 ATTCCAATGGAACTGGAGATATCATTTGGTGGCTGGAAACCATTTGGGATGGTTATTCT 1620

DB 1758 ATTCCAATGGAACTGGAGATATCATTTGGTGGCTGGAAACCATTTGGGATGGTTATTCT 1817

QY 1621 AAAGGTATCAACAGAAACTTTGGAAAACAGGCTTTATATCCCTCTCAAAAGTCCGAGAG 1680

DB 1818 AAAGGTATCAACAGAAACTTTGGAAAACAGGCGGCTATATCCCTCTCAAAAGTCCGAGAG 1877

QY 1681 AAGATAGAAACAGTCAAGTATCCCAATATCTCCTCAAGCTGAAATA 1727

DB 1878 AAGATAGAAACAGTCAAGTATCCCAATATCTCCTGAGGCTGAGAAATA 1924

## RESULT 2

US-09-442-629-9  
; Sequence 9, Application US/09442629  
; Patent No. 6291219

## GENERAL INFORMATION:

APPLICANT: TANIGUCHI, Naoyuki

SHIBA, Tetsuo

YANAGIDANI, Shusaku

TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Kenyon & Kenyon

STREET: 1025 Connecticut Avenue, N.W., Suite 600

CITY: Washington

STATE: DC

COUNTRY: US

ZIP: 20036

COMPUTER READABLE FORM:

MEDIUM TYPE: 3+ Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/442,629  
FILING DATE: 18-NOV-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/913,805A  
FILING DATE: 7 JAN 1998  
APPLICATION NUMBER: PCT/JP97/00171  
FILING DATE: 23 JAN 1997  
APPLICATION NUMBER: JP 192260  
FILING DATE: 22 JUL 1996  
APPLICATION NUMBER: JP 162813  
FILING DATE: 24 JUN 1996  
APPLICATION NUMBER: JP 161648  
FILING DATE: 21 JUN 1996  
APPLICATION NUMBER: JP 10365  
FILING DATE: 24 JAN 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Toffenetti, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 2356/3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2100 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 9:

US-09-442-629-9

Query Match 86.6%; Score 1496.6; DB 3; Length 2100;  
Best Local Similarity 91.7%; Pred. No. 0;  
Matches 1583; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY	1	ATGGGGGATGGAGCTGGTCTCTGGGTTGGATTATGCTCATCTCTTTTGGCTGGGGACC	60
DB	198	ATGGGGCCATGGAGCTGGTCTCTGGGTTGGATTATGCTCATCTCTTTTGGCTGGGGACC	257
QY	61	TTGTTATTTTATATAGTGGTTCATTGGTTCGAGATATGACCCCTTGATCACTCCAGC	120
DB	258	TTGCTGTTTATATAGTGGTTCATTGGTTCGAGATATGACCCCTTGATCACTCTAGC	317
QY	121	AGAGAACTCTCCAAAGTCTTGAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG	180
DB	318	CGAGAACTGTCCAAAGTCTGGCAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG	377
QY	181	AGGGAAATGGCTGAGTCTCTCGAATACAGAGGCCCATTTGACGGGGACAGCTACA	240
DB	378	AGGGAAATGGCCGAATCTCTCGGATACAGAGGCCCATTTGATCAGGGGCCAGCTATA	437
QY	241	GGAAGTCCGCTGTTTGAAGAACAGCTTGAAGGCCAAAGACAGATTGAAATTTAC	300
DB	438	GGAAGTACCGCTTGTGAAGAACAGCTTGTGAAGGCCAAAGACAGATTGAAATTTAC	497
QY	301	AAGAAACAGCTAGAAATGGTCTGGGGAAGGATCATGAAATCTTAAAGAGGAGATTGAA	360
DB	498	AAGAAACAGACCAAGAAATGGTCTGGGGAAGGATCATGAAATCTTGAAGAGGAGATTGAA	557
QY	361	AATGAGCTAAGAGCTCTGTTTCTTCAAGAGCGAACTGAAAGAAATTAAGCATTTA	420
DB	558	AATGAGCTAAGAGCTCTGTTTCTTCAAGAGCGAACTGAAAGAAATTAAGCACTTA	617
QY	421	GAGGAATGAATCCCAAGACATGCAGATGAATCTTTTGGATTTAGGCACCATGAA	480
DB	618	GAGGAATGAATCCCAAGACATGCAGATGAATCTTTTGGATTTAGGCACCATGAA	677
QY	481	AGGTCTATCATGACAGATCTATCTACCTCAGTCAACAGATGGAGCGAGGATGGCGT	540

DB	678	AGGTCTATAATGACGGATCTATATACTACTCCTCAGTCAGACAGATGGAGCGAGTGTGGCGG	737
QY	541	GAAGAGAGGCGCAAGATCTGACAGAGCTGTCTCAGCGGAGAAATAACATATCTCCAGAA	600
DB	738	GAAGAGAGGCGCAAGATCTGACAGAACTGGTTCAGCGGAGAAATAACATATCTCCAGAA	797
QY	601	CCTAAGAGCTGCAGCAAGCCAGGAAGCTGGTGTAAACATCAATAAAGGCTGTGGCTAT	660
DB	798	CCCAAGAGCTGCAGCAAGCCCAAGGCTGGTGTAAATATCAACAAAGGCTGTGGCTAT	857
QY	661	GGTTGTCAATCCCATCAGTGGTCTACTGTTTCATGATTGCTTATGGCACCAGGGAACA	720
DB	858	GGCTGTGAGTCTCATCATGTTGTTCTACTGTTTCATGATTGCTTATGGCACCAGGGAACA	917
QY	721	CTCATCTTGGAACTCTCAGAAATGGCGCTATCTACTGTTGATGGAGACTGTGTTTGA	780
DB	918	CTCATCTTGGAACTCTCAGAAATGGCGCTATGCTACTGTTGATGGAGACTGTATTTAGG	977
QY	781	CCTGTAAGTGAGACATCTACAGACAGATCTGGCTCTCCACTGGACACTGTGTCAGGTGAA	840
DB	978	CCTGTAAGTGAGACATCTACAGACAGATCTGGCATCTCCACTGGACACTGTGTCAGGTGAA	1037
QY	841	GTAATGACAAAAACATTCAGAGTGGTGGCTCCCAATTTGTAGACAGCTCCATCTCGG	900
DB	1038	GTGAAGGACAAAAATGTTCAAGTGGTGGCTCCCAATTTGTAGACAGCTCTTCATCCCGT	1097
QY	901	CCTCTTACTTACCACCTGGCTGTTCCAGAAAGCTTCAGACAGCTTCAGACAGCTCC	960
DB	1098	CCTCATATTTACCTTGGCTGTACCAAGAGACCTCGCAGATCGACTTGTAGAGTGCAT	1157
QY	961	GGTGACCTCGAGTGTGGTGGCTTCCAGTTTGTCAAATACTTTGATTCGTCCCAAGCT	1020
DB	1158	GGTGACCTCGAGTGTGGTGGCTTCCAGTTTGTCAAATACTTTGATTCGTCCCAAGCT	1217
QY	1021	TGGCTGGAAGGAAAAAGAGAGAGCCACCAAGAGCTTGGCTTCAACATCCAGTTAT	1080
DB	1218	TGGCTGGAAGGAAAAAGAGAGAGCCACCAAGAGCTTGGCTTCAACATCCAGTTAT	1277
QY	1081	GGAGTCCATGTGCAGCSCAGACAAAGTGGGACAGAAAGTGGGACAGAGCTTCCACCCATCGAG	1140
DB	1278	GGAGTCCATGTGCAGCSCAGACAAAGTGGGACAGAAAGTGGGACAGAGCTTCCACCCATCGAG	1337
QY	1141	GAGTACATGTGTACAGTTTGAAGAACATTTTTCAGCTTCTCGCACGAGAAATCCAAAGTGCAT	1200
DB	1338	GAGTACATGTGTGTGATGTTGAAGAACATTTTTCAGCTTCTCGCACGAGAAATCCAAAGTGCAT	1397
QY	1201	AAAAAAGAGTATATCTGGCTACTGATGATCTTCTTGTAAAGAGGAGGCAAGCAAG	1260
DB	1398	AAAAAAGAGTATATTTGGCCACAGATGACCCCTTCTTTTAAAGAGGAGGCAAGCAAG	1457
QY	1261	TACTCCAAATTAAGAAATTTATAGTGATACTCTATTTTCTTGGTTCAGCTGGATCACAAAT	1320
DB	1458	TACCCCAATTAAGAAATTTATAGTGATACTCTATTTTCTTGGTTCAGCTGGATCACAAAT	1517
QY	1321	CGGTACACAGAAATTCACCTTCGGGGTGTGATCTGATATACATTTTCTCTCACAGCT	1380
DB	1518	CGGTACACAGAAATTCACCTTCGGGGTGTGATCTGATATACATTTTCTCTCACAGCT	1577
QY	1381	GACTTCTAGTGTGTACTTTTTTCATCCAGCTCTGTCGGGTGCTTATGAAGATCATGCA	1440
DB	1578	GACTTCTAGTGTGTACTTTTTTCATCCAGCTCTGTCGAGTGTGCTTATGAAGATCATGCA	1637
QY	1441	ACCTGTGATCTGTGATGCTCTGGAACTTCGATCTTTTGGATGATCATCTACTATTTTGA	1500
DB	1638	ACACTATCTGTGATGCTCTGGAACTTCGATCTTTTGGATGATCATCTACTATTTTGG	1697
QY	1501	GGCCAAATCCCAAGTGTGTTTATCTCCACAAACCTTCAACCTGAACTGAAAGAGAA	1560
DB	1698	GGCCAAATCCCAAGTGTGTTTATCTCCACAAACCTTCAACCTGAACTGAAAGAGAA	1757
QY	1561	ATTCCAAATGGAACCTGGAGATATCATTTGGTGTGGCTGGAAACCATTTGGATGTTTCT	1620
DB	1758	ATTCCAAATGGAACCTGGAGATATCATTTGGTGTGGCTGGAAACCATTTGGATGTTTCT	1817

QY 1621 AAGAGTATCAACAGAAAACCTTGGAAAAACAGGCTTATATCCCTCTCAAAAGTCCGAGAG 1680  
DB 1818 AAGAGTGTCAACAGGAAATTTGGGAAGGACGGGCTTATATCCCTCTCAAAAGTCCGAGAG 1877  
QY 1681 AAGATAGAAAACAGTCAAGTATCCCATATCTCTGAAGCTGAAAAATA 1727  
DB 1878 AAGATAGAAAACGCTCAAGTATCCCATATCTCTGAAGCTGAAAAATA 1924

## RESULT 3

US-08-913-805A-1  
; Sequence 1, Application US/08913805A  
; Patent No. 6054304  
; GENERAL INFORMATION:  
; APPLICANT: TANIGUCHI, Naoyuki  
; APPLICANT: UOZUMI, Naofumi  
; APPLICANT: SHIBA, Tetsuo  
; APPLICANT: YANAGIDANI, Shusaku  
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3+ Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
; SOFTWARE: WordPerfect 6.1 Windows  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/913,805A  
; FILING DATE: 7 JAN 1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP97/00171  
; FILING DATE: 23 JAN 1997  
; APPLICATION NUMBER: JP 192260  
; FILING DATE: 22 JUL 1996  
; APPLICATION NUMBER: JP 162813  
; FILING DATE: 24 JUN 1996  
; APPLICATION NUMBER: JP 161648  
; FILING DATE: 21 JUN 1996  
; APPLICATION NUMBER: JP 10365  
; FILING DATE: 24 JAN 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toffenetti, Judith L.  
; REGISTRATION NUMBER: 39,048  
; REFERENCE/DOCKET NUMBER: 2356/3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-429-1776  
; TELEFAX: 202-429-0796  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1728 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-913-805A-1

Query Match 85.0%; Score 1469.4; DB 3; Length 1728;  
Best Local Similarity 90.7%; Pred. No. 0;  
Matches 1566; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 1 ATCGGGGATGACGTGGTTCCTGGCGTTGGATTATGCTCATTTCTTTTGGCCCTGGGGGACC 60  
DB 1 ATCGGGCCATGACGTGGTTCCTGGCGTTGGATTATGCTCATTTCTTTTGGCCCTGGGGGACC 60  
QY 61 TTGTTATTTTATATAGTGGTGCATTTGGTTTCGAGATAATGACCACCCCTGCATCTCCAGC 120

DB 61 TTGTTATTTTATATAGTGGTGCATTTGGTTTCGAGATAATGACCACCTCTGCATCTCTAGC 120  
QY 121 AGAGAACTCTCCAAGATTCTTGCAAGCTTGAAAGCTTTAAAAACAGCAAAATGAAGACTTG 180  
DB 121 CGAGAACTGTCCAAGATTCTTGCAAGCTTGAAAGCTTTAAAAACAAATAATGAAGACTTG 180  
QY 181 AGCGAAATGGCTGAGTCTCTCCGAATACCAAGAGGCCCAATTTGACCGGGGACAGCTACA 240  
DB 181 AGGAGAAATGGCTGAATCTCTCCGAATACCAAGAGGCCCAATTTGATCAGGGCCAGCTTCA 240  
QY 241 GGAAGAGTCCGTGTTTATAGAAACAGCTTGTAAAGGCCAAAGACAGATGAAATTTAC 300  
DB 241 GGAAGAGTCCGTGTTTATAGAAACAGCTTGTAAAGGCCCAAGACAGATGAAATTTAT 300  
QY 301 AAGAAACAAAGCTAGAAATGGTCTGGGAAGGATCATGAAATCTTAAAGAGGAGGATTGAA 360  
DB 301 AAGAAACAAACTAAATATGGTCCAGGGAAGGATCATGAAATCTTAAAGAGGAGGATTGAA 360  
QY 361 AATGGAGCTAAGAGCTCTGTTTTTTCTACAAAGCGAACTGAAAGAAATTTAAAGCAATTA 420  
DB 361 AATGGAGCTAAGAGCTCTGTTTTTTCTACAAAGTGAGTTGAAAGAAATTTAAAGCAATTA 420  
QY 421 GAAGGAAATGAATCTCAAAAGACATGCAGATGAAATCTTTTGGATTTTAGGACACCATGAA 480  
DB 421 GAAGGAAATGAATCTCAAAAGACATGCAGATGAAATCTTTTATCAGATTTGGGACATCATGAA 480  
QY 481 AGGTCTATCATGCAGATCTATCTACTCTCAAGTCAAAACAGATGAGAGGGGATTGGCGT 540  
DB 481 AGGTCTATCATGCAGATCTATCTACTCTCAAGTCAAAACAGATGAGAGGGGATTGGCGT 540  
QY 541 GAAAAGAGGCGCAAGAGATCTGACAGAGCTGTCAGCGGAGAAATAACATATCTCCAGAAAT 600  
DB 541 GAAAAGAGGCGCAAGAGATCTGACAGAGCTGTCAGCGGAGAAATAACATATCTCCAGAAAT 600  
QY 601 CCTAAGGACTGCAGAAAGCCAGGAAGCTGGTGTGTAAACATCAATTAAGGCTGTGGCTAT 660  
DB 601 CCTAAGGACTGCAGAAAGCCAGGAAGCTGGTGTGTAAATATCAACAAAGGCTGTGGCTAT 660  
QY 661 GGTGTCAACTCCATCAGCTGGTCTACTGTTTCTATGATTTGCTTATGGCACCAGCGAACA 720  
DB 661 GGTGTCAACTCCATCAGCTGGTCTACTGTTTCTATGATTTGCTTATGGCACCAGCGAACA 720  
QY 721 CTCATCTTTGGAATCTCAGAAATGGCGCTATGCTACTGCTGGATGGGAGACTGTGTTTAGA 780  
DB 721 CTCATCTTTGGAATCTCAGAAATGGCGCTATGCTACTGCTGGATGGGAGAACTGTGTTTAGA 780  
QY 781 CCTGTAAAGTGAACATGTACAGACAGATCTGGCCCTCTCCACTGGACACTGTGTAGTGA 840  
DB 781 CCTGTAAAGTGAACATGTACAGACAGATCTGGCAGCTCCACTGGACATTTGTCAGGTGAA 840  
QY 841 GTAAATGACAAAAACATTCAAGTGTGTGAGCTCCCAATTTGAGACAGCTCCATCTCTCGG 900  
DB 841 GTAAAGGACAAAAATGTTTCAGTGTGTTGAGCTCCCAATTTGAGACAGCTTCATCTCTCGT 900  
QY 901 CCTCTTTACTTTACCACTGGCTGTTTCCAGAGACCTTGCAGACCGACTCTTAAGAGTCCAT 960  
DB 901 CCTCCATATTTACCCCTGGCTGTCCAGAGACCTTGCAGATCGACTTGTACGAGTCCAT 960  
QY 961 GGTGACCTGTGAGTGTGGTGTCTCCAGATTTGTCAAAATCTTGTATCTGTCTCCACACCT 1020  
DB 961 GGTGATCTGTGAGTGTGGTGTCTCCAGATTTGTCAAGTACTTTGTTCGCGCCCAACCC 1020  
QY 1021 TGGCTGGAAGAGGAATAGAGAGCCACCAAGAGCTTTGGCTTCAAAATCTCCAGTTATT 1080  
DB 1021 TGGCTGGAAGAGGAATAGAGAGCCACCAAGAGCTTAGGCTTCAAAATCTCCAGTTATT 1080  
QY 1081 GGAGTCCATGTTCAGACGACAGACAAAAGTGGGAAACAGAGACGCTTCCACCCCATCGAG 1140  
DB 1081 GGAGTCCATGTTCAGACGACAGACAAAAGTGGGAAACAGAGACGCTTCCATCCCATCGAG 1140  
QY 1141 GAGTACATGTTGATACGTTTGAAGAAACNTTTTCAGCTTCTCGCACCGAGATCGAAGTGAAT 1200

Db 1141 GAATACACGGTGCACGTTGAAGAAGACTTTTCAGCTCTTTCGTCGAGAAATGCAAGTGGAT 1200  
Qy 1201 AAAAAAGAGTATATCTGCTACTGATGATCTGATCTGTTGTTAAAGGAGGCAAGCAAG 1260  
Db 1201 AAAAAAGGGTATTTGGCCACAGATGACCTGCTGTTGTTAAAGAGGCAAAACAAAG 1260  
Qy 1261 TACTCCAATATGAATTTATTAGTGATAACTCTATTCTTGGTCAGCTGGACTACACAAT 1320  
Db 1261 TACCCAGTTATGAATTTATTAGTGATAACTCTATTCTTGGTCAGCTGGACTACATAAT 1320  
Qy 1321 CGGTACACAGAAATTCATCTGGGGTGTGATCTGGGATATACATTTCTCTCAGAGCT 1380  
Db 1321 CGATATACAGAAATTCATCTGGGGTGTGATCTGGGATATACATTTCTCTCCAGGCA 1380  
Qy 1381 GACTTCTAGTGTGATCTTTTCATCCAGGTCCTCGGGTGTGATGATGAATCATGCAA 1440  
Db 1381 GACTTCTAGTGTGATCTTTTCATCCAGGTCCTCGGGTGTGATGATGAATCATGCAA 1440  
Qy 1441 ACCCTGCATCTGATCTCTGCGAACTTCCATCTTTGGATGACATCTACTATTTTGA 1500  
Db 1441 GCGCTGCATCTGATCTCTGCGAACTTCCATCTTTGGATGACATCTACTATTTTGA 1500  
Qy 1501 GGCACAAATGCCCAACATCAGATTCTGTTTATCTCAAACTCGAACTGAAGAGAA 1560  
Db 1501 GGCACAAATGCCCAACAAATTTCCATCTTATCTCAAACTCGAACTGAAGAGAA 1560  
Qy 1561 ATTCCAATGGAACTGGAGATATCATCTGTGTGGCTGGAACCAATGGGATGGTTATTCT 1620  
Db 1561 ATCCCATGGAACTGGAGATATTAATGTGTGGCTGGAATCATCTGGATGGCTATCTCT 1620  
Qy 1621 AAAGGTATCAACAGAAAACTTTGGAAAAACAGGCTTTATATCCCTCTCAAAAGTCGAGAG 1680  
Db 1621 AAAGGTATTAACAGAAAACTGGAGGAGCGGCTATATCCCTCTCAAAAGTCGAGAG 1680  
Qy 1681 AAGATAGAAACAGTCAAGTATCCCAATATCTCTGAACTGAAATA 1727  
Db 1681 AAGATAGAAACAGTCAAGTATCCCAATATCTCTGAACTGAAATA 1727

RESULT 4  
US-09-442-629-1  
; Sequence 1, Application US/09442629  
; Patent No. 6291219  
; GENERAL INFORMATION:  
; APPLICANT: TANIGUCHI, Naoyuki  
; UOZUMI, Naofumi  
; SHIBA, Tetsuo  
; YANAGIDANI, Shusaku  
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3+ Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
; SOFTWARE: WordPerfect 6.1 Windows  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/442,629  
; FILING DATE: 18-NO. 6291219-1999  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/913,805A  
; FILING DATE: 7 JAN 1998  
; APPLICATION NUMBER: PCT/JP97/00171  
; FILING DATE: 23 JAN 1997  
; APPLICATION NUMBER: JP 192260  
; FILING DATE: 22 JUL 1996  
; APPLICATION NUMBER: JP 162813

; FILING DATE: 24 JUN 1996  
; APPLICATION NUMBER: JP 161648  
; FILING DATE: 21 JUN 1996  
; APPLICATION NUMBER: JP 10365  
; FILING DATE: 24 JAN 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toffenetti, Judith L.  
; REGISTRATION NUMBER: 39,048  
; REFERENCE/DOCKET NUMBER: 2356/3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-439-1776  
; TELEFAX: 202-429-0796  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1728 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-442-629-1  
  
Query Match 85.0%; Score 1469.4; DB 3; Length 1728;  
Best Local Similarity 90.7%; Pred. No. 0;  
Matches 1566; Conservative 0; Mismatches 161; Indels 0; Gaps 0;  
  
Qy 1 ATCGGGGCGATGAGCTGTTCTCTGGCGTTGGATTATGCTCATTTCTTTTGGCTGGGGACC 60  
Db 1 ATCGGGGCGATGAGCTGTTCTCTGGCGTTGGATTATGCTCATTTCTTTTGGCTGGGGACC 60  
  
Qy 61 TTGTTATTTTATATAGTGGTTCATTTGGTTTCGAGATAATGACCCCTGATCATCTCCAGC 120  
Db 61 TTGCTATTTTACATAGTGGTTCATTTGGTTTCGAGATAATGACCCCTGATCATCTCTAGC 120  
  
Qy 121 AGAGAACTCTCCAAAGATCTTGGAAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 180  
Db 121 CGAGAACTGTCCAAAGATTTTGGCAAGCTGGAACGCTTAAACAAACAAATGAAGACTTG 180  
  
Qy 181 AGGCGAATGGCTGAGTCTCTCCGAATACCAAGAGCCCTATTGACAGGGGACAGCTACA 240  
Db 181 AGGAGATGGCTGTAATCTCTCCGAATACCAAGAGCCCTATTGATGAGGGGACAGCTTCA 240  
  
Qy 241 GGAAGAGTCCGTGTTTATGAAGAAACAGCTTGTATAGGCCCAAAAGAACAGATTGAAAAT 300  
Db 241 GGAAGAGTCCGTGCTTTATGAAGAGCAATTTATGAAGGCCAAAGAACAGATTGAAAAT 300  
  
Qy 301 AAGAAACAGCTAGAAATGGTCTCGGGAAGAGATCATGAATCTTAAAGAGAGAGATTGAA 360  
Db 301 AAGAAACAAACTAAAAATGGTCTCCAGGGAAGAGATCATGAATCTTAAAGAGAGAGATTGAA 360  
  
Qy 361 AATGAGCTAAAGAGCTCTGTTTCTCAAAAGCGAAGCTGAAAGAAATTAAGCAATTAA 420  
Db 361 AATGAGCTAAAGAGCTCTGTTTCTCAAAAGCTGAGTTTGAAGAAATTAAGCAATTAA 420  
  
Qy 421 GAAGCAATGAATCTCCAAAGACATGCAATGAAATTTCTTTGGATTATAGGACACCATGAA 480  
Db 421 GAAGCAATGAATCTCCAAAGACATGCAATGAAATTTCTATCAGATTGGGACATCATGAA 480  
  
Qy 481 AGGTCTATCATGACAGATCTATATCTAGTCAAAAGAGATGAGAGCGGGGATTTGGCT 540  
Db 481 AGGTCTATATGACGGATCTATATCTAGTCAAAAGAGATGAGAGCGGGGATTTGGCT 540  
  
Qy 541 GAAAGAGGCGCAAGATCTGACAGAGCTGGTCAGCGGAGAAATAACATATCTCCAGAT 600  
Db 541 GAAAGAGGCGCAAAAGATCTGACAGAGCTGGTCAGCGGAGAAATAACATATCTCCAGAT 600  
  
Qy 601 CCTAAGGACTCGCAAAAGCCAGGAAGCTGGTGTGTAACATCAATAAAGGCTGTGGCTAT 660  
Db 601 CCCAAGGACTCGCAAAAGCCAGGAAGCTAGTGTGTAATATCAACAAAGGCTGTGGCTAT 660  
  
Qy 661 GGTTGTCAACTCCATGAGCTGGTGTACTGTTTCTATGATTGCTTATGGCACCCAGCAACA 720  
Db 661 GGCTGTGAGCTCCATCATGTAGTGTACTGCTTTTATGATTGCTATGGCACCCAGCAACA 720









## TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-721-8200  
TELEFAX: 202-721-8250  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2116  
TYPE: Nucleic acid  
STRANDEDNESS: Double  
TOPOLOGY: Linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Human  
US-09-000-041A-1

Query Match 2.2%; Score 37.6; DB 3; Length 2116;  
Best Local Similarity 49.6%; Pred. No. 0.18;  
Matches 126; Conservative 0; Mismatches 124; Indels 4; Gaps 1;  
QY 280 AAAGAACAGATTGAAATTTACAAGAAACAAAGCTAGAAATGCTCTGGGAAGGATCATGAA 339  
DB 1507 AATGAACAATGACCTCGCTATCCCAAGCCATCAGCTCTGGAAGGGATCCCTGA 1566  
QY 340 ATCTTAAAGAGAGGATTGAAATTAAGAGCTAAAGAGCTCTGTTTTTTTCTACAAAGCGAA 399  
DB 1567 ATCTCTCAGGAGCATTGTACACAAAGAAATGGCTTTAGCTATTCTTACAAAGGAAA 1626  
QY 400 CT----GAAGAAATTAAGCATTTAGAGGAAATGAATCCCAAGACATGAGATGAAT 455  
DB 1627 GGAGTATTGGAATTTCAACAAACCATGATCTCAAGGTAGAACCTGGACATCCCAAGATCCAT 1686  
QY 456 TCTTTTGATTTAGGACACCATGAAAGGTCTATCATGACAGATCATACTACCTCAGTCA 515  
DB 1687 CCTAAGATTTTATGGCTGTGATGGACCAACAGACAGAGTTAAGAGGACACAGCCC 1746  
QY 516 AACAGATGGAGCAG 529  
DB 1747 ACCAGATGATGAG 1760

## RESULT 9

US-09-134-001C-2083  
Sequence 2083, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
PRIOR FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 2083  
LENGTH: 1596  
TYPE: DNA  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-2083

Query Match 2.2%; Score 37.2; DB 4; Length 1596;  
Best Local Similarity 63.3%; Pred. No. 0.2;  
Matches 57; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
QY 237 TACAGAGAGTCCTGTTTAAAGAAACAGCTTCTTAAGGCCAAGAACAGATTGAAAA 296  
DB 177 TAAAGAGCAGCAATATTAAGAAAGAAATTAATCTTGAAGGCAAGAGAGAACCAAT 236  
QY 297 TTACAGAAACAAAGCTAGAAATGGTCTGGG 326  
DB 237 TTAAAGAGACAGCTGAGATGACTGCG 266

## RESULT 10

US-08-933-750C-54/c  
Sequence 54, Application US/089333750C  
Patent No. 5932442  
GENERAL INFORMATION:  
APPLICANT: Lal, Preeti  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Bandman, Olga  
APPLICANT: Shah, Purvi  
APPLICANT: Au-Young, Janice  
APPLICANT: Yue, Henry  
APPLICANT: Guegler, Karl J.  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES  
NUMBER OF SEQUENCES: 98  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/933,750C  
FILING DATE: September 23, 1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0356 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 54:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1750 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: HMCINOT01  
CLONE: 9476  
US-08-933-750C-54

Query Match 2.1%; Score 36.8; DB 2; Length 1750;  
Best Local Similarity 54.4%; Pred. No. 0.29;  
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;  
QY 304 AAACAAGCTAGAAATGTCTCTGGGAAGGATCATGAAATCTTTAAGAGGAGGATTGAAAA 363  
DB 1718 AAAAATCTAGAGAGAGAGAGAGAGAGATCAAAATTAACCTATGAAGATTTATAGAAA 1659  
QY 364 GGAGCTAAGAGCTCTGTTTTTTTCTCAAGCCGACTGAAGAAATTAAGCATTTAGAA 423  
DB 1658 ACACGAGTAGGAAGGGTTCTTTCAAGCAACCAACAGACAGTTACGATAACTTAAA 1599  
QY 424 GGAATGACTCCAAA 439  
DB 1598 AGAAATGCTTTCCCA 1583

## RESULT 11









Db 61 TTGTTATTTATATAGTGGTCAATTTGGTTCGAGATAATGACCCCTGATCACTCCAGC 120  
Qy 121 AGAGAACTCTCAAGATCTTCCAAAGCTTGAACGCTTAAACACACAGAAAATGAAGACTTG 180  
Db 121 AGAGAACTCTCCAGATCTTCCAAAGCTTGAACGCTTAAACACACAGAAAATGAAGACTTG 180  
Qy 181 AGGCGAATGGCTGAGTCTCTCCGAATACCCAGAAAGCCCAATTTGACAGGGACAGCTACA 240  
Db 181 AGGCGAATGGCTGAGTCTCTCCGAATACCCAGAAAGCCCAATTTGACAGGGACAGCTACA 240  
Qy 241 GGAAGAGTCCTGTTTGAAGAACAGACTTGTAAAGGCGCAAGAAACAGATGAAATTTAC 300  
Db 241 GGAAGAGTCCTGTTTGAAGAACAGACTTGTAAAGGCGCAAGAAACAGATGAAATTTAC 300  
Qy 301 AAGAAACAGCTAGAAATGGTCTGGGGAAGGATCATGAAATCTTAAAGAGGAGGATGAA 360  
Db 301 AAGAAACAGCTAGAAATGGTCTGGGGAAGGATCATGAAATCTTAAAGAGGAGGATGAA 360  
Qy 361 AATGAGCTTAAAGAGCTCTGGTCTTCTCAAAAGCGAACTGAAGAAATTTAAAGCATTTA 420  
Db 361 AATGAGCTTAAAGAGCTCTGGTCTTCTCAAAAGCGAACTGAAGAAATTTAAAGCATTTA 420  
Qy 421 GAAGGAATGAATCTCAAGAGCATGAGATGAAATCTTTTGGATTTAGGACACCATGAA 480  
Db 421 GAAGGAATGAATCTCAAGAGCATGAGATGAAATCTTTTGGATTTAGGACACCATGAA 480  
Qy 481 AGGCTTATCATGACAGATCTACTACCTCAGTCAACAGATGAGAGGAGGATTTGGCGT 540  
Db 481 AGGCTTATCATGACAGATCTACTACCTCAGTCAACAGATGAGAGGAGGATTTGGCGT 540  
Qy 541 GAAAGAGGCGCAAGATCTGACAGAGTGGTCCAGCGGAGAAATTAACATATCTCCAGAA 600  
Db 541 GAAAGAGGCGCAAGATCTGACAGAGTGGTCCAGCGGAGAAATTAACATATCTCCAGAA 600  
Qy 601 CTAAGAGCTGAGCAAGACCCAGGAAGTGGTGTAAATCAATGAAAGGCTGGGCTAT 660  
Db 601 CTAAGAGCTGAGCAAGACCCAGGAAGTGGTGTAAATCAATGAAAGGCTGGGCTAT 660  
Qy 661 GGTGCTCACTCACTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
Db 661 GGTGCTCACTCACTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
Qy 721 CTCACTCTGGAATCTCAGAAATTTGGCGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780  
Db 721 CTCACTCTGGAATCTCAGAAATTTGGCGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780  
Qy 781 CCTGTAAGTGAGACATGTACAGACAGATCTGGCCCTCTCCACTGGGACACTGGTCAAGTGA 840  
Db 781 CCTGTAAGTGAGACATGTACAGACAGATCTGGCCCTCTCCACTGGGACACTGGTCAAGTGA 840  
Qy 841 GTAAATGACAAACATTCAGTGGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
Db 841 GTAAATGACAAACATTCAGTGGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
Qy 901 CCTCCTCTTACTTACCACTGGCTGCTCCAGAAAGCTTGGCAGACCGACTCTTAAGAGTCCAT 960  
Db 901 CCTCCTCTTACTTACCACTGGCTGCTCCAGAAAGCTTGGCAGACCGACTCTTAAGAGTCCAT 960  
Qy 961 GGTGACCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1020  
Db 961 GGTGACCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1020  
Qy 1021 TGGCTGGAAGGAATGAGAAAGCCACCAAGAGCTTGGCTTCAACATCCAGTTAT 1080  
Db 1021 TGGCTGGAAGGAATGAGAAAGCCACCAAGAGCTTGGCTTCAACATCCAGTTAT 1080  
Qy 1081 GGAGTCCATGTGACAGCGACACAGAAAGTGGGAAACAGAGAGCTTCCACCCCATCGAG 1140  
Db 1081 GGAGTCCATGTGACAGCGACACAGAAAGTGGGAAACAGAGAGCTTCCACCCCATCGAG 1140  
Qy 1141 GAGTACATGGTACAGTGTGAAGAAATTTTCAAGTCTTCGACGAGAAATGCAAGTGGAT 1200

Db 1141 GAGTACATGGTACACGCTTGAAGAAATTTTTCAGCTTCTCGCACGCAAGATGCAAGTGGAT 1200  
Qy 1201 AAAAAAGAGATATATCTGGCTACTGATGATCTCTACTTTGTTTAAAGGAGGCAAAAGCAAAG 1260  
Db 1201 AAAAAAGAGATATATCTGGCTACTGATGATCTCTACTTTGTTTAAAGGAGGCAAAAGCAAAG 1260  
Qy 1261 TACTCCAAATTAATGAATTTATAGTGAATACTCTATTTCTTGGTCACTGAGTCAACAAAT 1320  
Db 1261 TACTCCAAATTAATGAATTTATAGTGAATACTCTATTTCTTGGTCACTGAGTCAACAAAT 1320  
Qy 1321 CGGTACACAGAAAATTCACCTTCGGGGTGTGATCTGATGATATACACTTTCTCTCACAGGCT 1380  
Db 1321 CGGTACACAGAAAATTCACCTTCGGGGTGTGATCTGATGATATACACTTTCTCTCACAGGCT 1380  
Qy 1381 GACTTTCTAGTGTACTTTTTTCATCCAGGTCTGTCTGGGTGCTTATGAAATCATGCAA 1440  
Db 1381 GACTTTCTAGTGTACTTTTTTCATCCAGGTCTGTCTGGGTGCTTATGAAATCATGCAA 1440  
Qy 1441 ACCGTGATCTGATGCTCTCGAACTTCGAACTTCGAACTTCGAACTTCGAACTTCGAACTTCGAA 1500  
Db 1441 ACCGTGATCTGATGCTCTCGAACTTCGAACTTCGAACTTCGAACTTCGAACTTCGAACTTCGAA 1500  
Qy 1501 GGCCAAAATGCCCAATTCAGATTTGCTTTTATCTCACAACCTCGAACTGAAAGAGAA 1560  
Db 1501 GGCCAAAATGCCCAATTCAGATTTGCTTTTATCTCACAACCTCGAACTGAAAGAGAA 1560  
Qy 1561 ATTCCAATGGAACCTGGAGATATCATTTGTTGCTGGAAACCAATTTGGATGGTTATTTCT 1620  
Db 1561 ATTCCAATGGAACCTGGAGATATCATTTGTTGCTGGAAACCAATTTGGATGGTTATTTCT 1620  
Qy 1621 AAAGGTATCAACAGAAAATTTGAAAACAGAGCTTATATCCCTCTCAAAAGTCCGAGAG 1680  
Db 1621 AAAGGTATCAACAGAAAATTTGAAAACAGAGCTTATATCCCTCTCAAAAGTCCGAGAG 1680  
Qy 1681 AAGATAGAAACAGTCAAGTATCCACATATCTCGAAGCTGAAAAATAG 1728  
Db 1681 AAGATAGAAACAGTCAAGTATCCACATATCTCGAAGCTGAAAAATAG 1728

## RESULT 2

US-09-971-773-1  
; Sequence 1, Application US/09971773  
; Publication No. US20030115614A1  
; GENERAL INFORMATION:  
; APPLICANT: Yutaka KANDA  
; APPLICANT: Mitsuo SATOH  
; APPLICANT: Kazuyasu NAKAMURA  
; APPLICANT: Kazuhisa UCHIDA  
; APPLICANT: Toyohide SHINKAWA  
; APPLICANT: Naoko YAMANE  
; APPLICANT: Motoo YAMASAKI  
; APPLICANT: No. US20030115614A1uo HANA1  
; TITLE OF INVENTION: ANTIBODY COMPOSITION-PRODUCING CELL  
; FILE REFERENCE: 249-202  
; CURRENT APPLICATION NUMBER: US/09/971.773  
; PRIOR FILING DATE: 2002-08-30  
; PRIOR APPLICATION NUMBER: JP 2000-308526  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 60/268,926  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 2008  
; TYPE: DNA  
; ORGANISM: Cricetulus griseus  
US-09-971-773-1

Query Match: 91.7%; Score 1584; DB 11; Length 2008;  
Best Local Similarity 94.8%; Pred. No. 0;  
Matches 1638; Conservative 0; Mismatches 90; Indels 0; Gaps 0;  
Qy 1 ATCGGGGATGGACTGGTTCCTGCGTGGATTATGCTCATCTTTTGGCTGGGGGACC 60

Db 100 ATCGGGGACGACTGGTCTCTGGCGTTGGATTATGCTCAATCTTTTGGCCCTGGGGACC 159  
Qy 61 TTGTTATTTTATATAGGTGGTCATTTGGTTCGAGATAATGACCACCTTGATCACTCCAGC 120  
Db 160 TTTATGTTTATATAGGTGGTCATTTGGTTCGAGATAATGACCACCTTGATCACTCCAGC 219  
Qy 121 AGAGAACTCTCCAGATTTCTGCAAGCTTGAAACGCTTTAAACAGCAAAATGAAGACTTG 180  
Db 220 AGAGAACTCTCCAGATTTCTGCAAGCTTGAAACGCTTTAAACAGCAAAATGAAGACTTG 279  
Qy 181 AGGGAATGGCTGAGTCTCTCGAATACAGAAAGCCCCATGACCGAGGAGACAGCTACA 240  
Db 280 AGGGAATGGCTGAGTCTCTCGAATACAGAAAGCCCCATGATGATGACCGAGGAGACAGTACA 339  
Qy 241 GGAAGAGTCCGTGTTTAAAGAAACAGCTTGTTAAGGCAAGAAACAGATTTGAATAATAC 300  
Db 340 GGAAGAGTCCGTGTTTAAAGAAACAGCTTGTTAAGGCAAGAAACAGATTTGAATAATAC 399  
Qy 301 AAGAAACAAAGCTAGAAATGGTCTGGGGAAGGATCATGAAATCTTTAAGAAAGGAGGATTCGA 360  
Db 400 AAGAAACAAAGCTAGAAATGGTCTGGGGAAGGATCATGAAATCTTTAAGAAAGGAGGATTCGA 459  
Qy 361 AATGAGGCTAAGAGCTCTGTTTCTTCTAAGAGCAACTGAAGAAATTAAGGATTTA 420  
Db 460 AATGAGGCTAAGAGCTCTGTTTCTTCTAAGAGCAACTGAAGAAATTAAGGATTTA 519  
Qy 421 GAAGAAATGACCTCAAGACATGACAGATGAAATCTTTTGGATTTAGGACACCATGAA 480  
Db 520 GAAGAAATGACCTCAAGACATGACAGATGAAATCTTTTGGATTTAGGACATCATGAA 579  
Qy 481 AGGTCTATCATGACAGATCTATCTACCTCAAGACATGAGAGAGGAGGATTTGGCGT 540  
Db 580 AGGTCTATCATGACAGATCTATCTACCTCAAGACATGAGAGAGGAGGATTTGGCGG 639  
Qy 541 GAAAGAGGCAAGATCTGACAGAGTGGTCCAGCGGAGAAATTAACATATCTCCAGAA 600  
Db 640 GAAAGAGGCAAGATCTGACAGAGTGGTCCAGCGGAGAAATTAACATATCTCCAGAA 699  
Qy 601 CCTAAGGAGTGCAGAAAGCCAGGAGGCTGGTGTATCATCAATTAAGGCTGGCTAT 660  
Db 700 CCTAAGGAGTGCAGAAAGCCAGGAGGCTGGTGTATCATCAATTAAGGCTGGCTAT 759  
Qy 661 GTTCTCAACTCCATCACTGCTGCTACTGTTTCTATGTTTCTATGTCACCCAGCGAACA 720  
Db 760 GGATGTCACCTCCATCATGCTGTTTCTGTTCTCATGTTGTTATGTCACCCAGCGAACA 819  
Qy 721 CTCATCTTTGGAATCTCAGAAATGGCGCTATGCTACTGTTGGATGGGAGCTGTGTTTGA 780  
Db 820 CTCATCTTTGGAATCTCAGAAATGGCGCTATGCTACTGTTGGATGGGAGCTGTGTTTGA 879  
Qy 781 CCTGTAAGTGAGACATGTACAGACAGATCTGGCCCTCTCCATGGACACTGTCCAGTGAA 840  
Db 880 CCTGTAAGTGAGACATGTACAGACAGATCTGGCCCTCTCCATGGACACTGTCCAGTGAA 939  
Qy 841 GTAATGACAAAAACATTCAGTGTGCTGAGTCTCCCATTTGATAGACAGCTCCATCTCGG 900  
Db 940 GTGAAGGACAAAATGTTCAAGTGTGCTGAGTCTCCCATTTGATAGACAGCTCCATCTCGT 999  
Qy 901 CCTCTTACTTACCATCTGGCTGTTCCAGAAAGACCTTTGACAGACCGACTCTTAAGAGTCCAT 960  
Db 1000 CCTCTTACTTACCATCTGGCTGTACAGAAAGACCTTTGACAGATCGACTCTCGAGAGTCCAT 1059  
Qy 961 GGTGACCTCGAGTGTGGTGTGCTCCAGTTTGTCAAAATCTTGATTCGTCACAACCT 1020  
Db 1060 GGTGATCTCGAGTGTGGTGTGCTCCAGTTTGTCAAAATCTTGATTCGTCACAACCT 1119  
Qy 1021 TGGCTGGAAGGAAATAGAGAGACCCAGAGAGCTTTGGCTCAACATCCAGTTAT 1080  
Db 1120 TGGCTGGAAGGAAATAGAGAGACCCAGAGAGCTTTGGCTCAACATCCAGTTAT 1179  
Qy 1081 GGAGTCCATGTACAGACGACACAGAAAGTGGGAAACAGAGAGCTTTCCACCCCATCCAG 1140

Db 1180 GGATCCATGTACAGACGACCTGACAAAAGTGGGAAACAGAAAGCAGCTTCCATCCCATGAG 1239  
Qy 1141 GAGTACATGGTACACGTTGAAGAAACATTTTTCAGCTTCTCGCACGACAGAAATGCAAGTGGAT 1200  
Db 1240 GAATACATGGTACACGTTGAAGAAACATTTTTCAGCTTCTCGAACGACAGAAATGCAAGTGGAT 1299  
Qy 1201 AAAAAAGAGATATATCTGGCTACATGATCCTACTTTTAAAGAGGCAAGAAAG 1260  
Db 1300 AAAAAAGAGATATCTGGCCACTGATGACCCCTTCTTTGTTAAAGAGGCAAGAAAG 1359  
Qy 1261 TACTCCAAATATGAATTTATAGTGAATACTCTATTTTCTGGTCAGCTGGATCACAAAT 1320  
Db 1360 TACTCCAAATATGAATTTATAGTGAATACTCTATTTTCTGGTCAGCTGGATCACAAAT 1419  
Qy 1321 CGGTACACAGAAAATTCATCTCGGGGTGTGATCTCGGATATACATTTTCTCTCACAGGCT 1380  
Db 1420 CGATACACAGAAAATTCATCTCGGGGTGTGATCTCGGATATACATTTTCTCTCCAGGCT 1479  
Qy 1381 GACTTTCTAGTGTGTAATTTTTCATCCAGGCTGTGTGGGTGCTTTATGAAATCATGCAA 1440  
Db 1480 GACTTCTGTGTGTAATTTTTCATCCAGGCTGTGTAGGTTGCTTTATGAAATCATGCAA 1539  
Qy 1441 ACCCTGATCTGATGCTCTCGAACTTCCATTTCTTTGGATGACATCTACTATTTTGA 1500  
Db 1540 AACTGATCTGATGCTCTCGAACTTCCATTTCTTTAGATGACATCTACTATTTTGA 1599  
Qy 1501 GGCCAAAATGCCAATCAGATTTGCTTTTATCTCACAACTCTGAACTGAAAGAGAA 1560  
Db 1600 GGCCAAAATGCCAATCAGATTTGCTTTTATCTCACAACTCTGAACTGAAAGAGAA 1659  
Qy 1561 ATTCCAATGGAACCTGGAGATATCATTTGGTGTGCTGGAAACCATTTGGATGTTTATCT 1620  
Db 1660 ATCCCATGGAACCTGGAGATATCATTTGGTGTGCTGGAAACCATTTGGAATGTTTATCT 1719  
Qy 1621 AAAGGTATCAACAGAAACCTTTGGAAACAGGCTTATATCCCTCTCAAAAGTCCGAGAG 1680  
Db 1720 AAAGGTATCAACAGAAACCTTTGGAAACAGGCTTATATCCCTCTCAAAAGTCCGAGAG 1779  
Qy 1681 AAGATAGAAACAGTCAAGTATCCACATATCTCGAAGCTGAAATAG 1728  
Db 1780 AAGATAGAAACAGTCAAAATACCTACATATCTCGAAGCTGAAATAG 1827

RESULT 3  
US-09-839-136-9  
; Sequence 9, Application US/09839136  
; Patent No. US20020081694A1  
; GENERAL INFORMATION:  
; APPLICANT: Naoyuki TANIGUCHI et al.  
; TITLE OF INVENTION: ALPHA 1-6 FUCOSYLTRANSFERASE  
; FILE REFERENCE: 2356-7  
; CURRENT APPLICATION NUMBER: US/09/839,136  
; CURRENT FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: 09/442,629  
; PRIOR FILING DATE: 1999-11-18  
; PRIOR APPLICATION NUMBER: 08/913,805  
; PRIOR FILING DATE: 1998-01-07  
; PRIOR APPLICATION NUMBER: PCT/JP97/00171  
; PRIOR FILING DATE: 1997-01-23  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 2100  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (198)....(1925)  
US-09-839-136-9

Query Match 86.7%; Score 1498.2; DB 9; Length 2100;  
Best Local Similarity 91.7%; Pred. No. 0;  
Matches 1584; Conservative 0; Mismatches 143; Indels 0; Gaps 0;





```

; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (3002)..(3002)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-381

```

Query Match 86.2%; Score 1490; DB 15; Length 3007;  
Best Local Similarity 91.7%; Pred. No. 0;  
Matches 1585; Conservative 1; Mismatches 141; Indels 1;

Qy	1	ATCGGGCATGGACTGGTTCCCTGGCGTTTGGATATATGCTCAATTCCTTTTGGCTCGCTGGGGGACC	60
Db	420	ATCGGGCCATGGACTGGTTCCCTGGCGTTGGAAATATGCTCAATTCCTTTTGGCTCGGGGGACC	479
Qy	61	TTGCTTATTTTATATAGTGTGCTCAATTTGGTTTCGAGATAATGACCAACCTGATCACTCCAGC	120
Db	480	TTGCTGTTTTTATATAGTGTGCTCACTTGGTACGAGATAATGACCATCTGATCACTCTAGC	539
Qy	121	AGAGAACTCTCCAAGATCTTTGCAAAAGCTTGAACGCTTTAAAAACAGCAAAATTAAGAACTTGG	180
Db	540	CGAGAACTGTCCAAGATCTCTGGCAAAAGCTTGAACGCTTTAAAAACAGCAGAAATGAAGAACTTG	599
Qy	181	AGSGGAATGCTG-AGTCTCTCCGATACAGAAAGGCCCCCATTTGACAGGGGACAGCTAC	239
Db	600	AGSGGAATGCGCAAAATCTCTCCGGATACAGAAAGGCCCTATTGATCAGGGGGCCAGCTAT	659
Qy	240	AGGAAGAGTCCGTGTTTTAGAGAAACAGCTTGTTTAAGGCCCAAGAAACAGAAATTCAAAATTA	299
Db	660	AGGAAGATGACGGTTTTTAGAGAGACAGCTTGTTTAAGGCCCAAGAAACAGAAATTCAAAATTA	719
Qy	300	CAAGAAACAAAGCTAGAAATCGTCTGGGGAAAGGATCATGAAATCTTAAGAAAGGAGGATTTGA	359
Db	720	CAAGAAACAGCACAGAAATGCTCTGGGGAAAGGATCATGAAATCTCTGAGGAGGAGGATTTGA	779
Qy	360	AAATGGAGCTAAGAGAGCTCTGGTTTTTTCTACAAAGCGAACTGAGAGAAATTAAGACATTT	419
Db	780	AAATGGAGCTAAGAGAGCTCTGGTTTTTTCTACAGAGTGAATTAAGAGAAATTAAGAGAACTT	839
Qy	420	AGAAGGAAATGAACCTCCAAAGACATGCAGATGAAATTCCTTTGGATTTAGACACACATGA	479
Db	840	AGAAGGAAATGAACCTCCAAAGACATGCAGATGAAATTCCTTTGGATTTAGACATCATGA	899
Qy	480	AAGSTCTPATGACAGATCTTACTACTCTCAGTCAAAACAGATGGAGCAGGGGATTTGGCG	539
Db	900	AAGSTCTPATGACGGATCTTACTACTCTCAGTCAAGATGGAGCAGGTGATTTGGCG	959
Qy	540	TGAANAAGAGGCCAAAGATCTGACAGAGCTGGTCCAGCGGAGAAATAACATATCTCCAGAA	599
Db	960	GGAAAAAGAGGCCAAAGATCTGACAGAACTGGTTTCAGCGGAGAAATAACATATCTTCAGAA	1019
Qy	600	TCCTTAAGSACTTGACAGCAAGCCAGGAAGCTGGTGTGTAAATCAATCAATAAAGCGCTGTGGCTA	659
Db	1020	TCCCAAGSACTTGACAGCAAGCCAAAGCTGGTGTGTAAATCAACAAAGCTGTGGCTA	1079
Qy	660	TGGTTGTCAATCCATCCATCCAGTGGTCTACTGTCTTCATGATTCCTTAATGGCACCCAGCGAAC	719
Db	1080	TGGCTGTCACTCCATCATGTGTGTCTACTGCTTCATGATTCGATATGGCACCCAGCGAAC	1139
Qy	720	ACTCATCTTGGAACTCAGAAATTTGGCGCTATGCTACTGGTGGATGGGAGACTGTGTTTAG	779
Db	1140	ACTCATCTTGGAACTCAGAAATTTGGCGCTATGCTACTGGTGGATGGGAGACTGTATTTAG	1199
Qy	780	ACCTGTAAGTGAGACATGTACAGACAGATCTGGCGCTCCACCTGGACACATGGTCAAGTGA	839
Db	1200	GCCTGTAAAGTGAGACATGTACAGACAGATCTGGCGATCTCCACTGGACACATGGTCAAGTGA	1259
Qy	840	AGTAAATGACAAAAACAAATTCAGTGGTTCGAGCTCCCAATTTGTAGACAGCCCTCCATCTCTCG	899
Db	1260	AGTGAAGACAAAAATGTTCAAGTGGTTCGAGCTCCCAATTTGTAGACAGCTCTTCATCCCCG	1319
Qy	900	GCCTCCTTTACTTACCACTGGCTGTTTCAGAAAGACTTTGCAGACCGACTCTCTAAGAGTCCA	959
Db	1320	TCCTCCATATTTACCTTTGGCTGTACAGAAAGACTTCGCAGATCGAATTTGTACAGTGA	1379

	960	TGTTGACCCCTGCAGTGTGGTGCGGTGCCAGTTTGTCAAAATACTTGAATTCGTCCACAACC	1019
Qy			
Db	1380	TGTTGACCCCTGCAGTGTGGTGCGGTGTCTCAGTTTGTCAAATACTTGTATCGGCCACAGCC	1439
Qy	1020	TTGGCTGGAAAAGGAANTAGAAGAAGCCACCAAGAAGCTTGGCTTCAAAACATCCAGTTAT	1079
Db	1440	TTGGCTTAGAAAAAAGAANAATAGAGAAGGCCAACAGAAGCTTGGCTTCAAAACATCCAGTTAT	1499
Qy	1080	TGGAGTCCATGTTCAGAA'GCAACAGAAAGTGGGAA'CAGAAGCAGCTTCCACCCCCATCGA	1139
Db	1500	TGGAGTCCATGTTCAGAA'GCAACAGAAAGTGGGAA'CAGAAGCTTCCATCCCATTGA	1559
Qy	1140	GGAAGTACATGGTACACGTTTGAAGAAACATTTTTCAGCTTCTCGCAGCGCAGAA'TGCAAGTGGGA	1199
Db	1560	AGAGTACATGGTGCATGTTTGAAGAAACATTTTTCAGCTTCTTGCACGCGAGAA'TGCAAGTGGGA	1619
Qy	1200	TAAAAAAGAGTATATCTGGCTACTGATGATCCTACTTTGTAAAGAGGACAAAGACAAA	1259
Db	1620	CAAAAAAGAGTGTATTTGGCCACAGATGACCTTCTTTATTAAGGAGGCAAAACAAA	1679
Qy	1260	GTACTCCAATPATGAATTTATTAGTGATAACTCTATTTTCTTGGTCAGCTGGACTCACAAA	1319
Db	1680	GTACCCCAATPATGAATTTATTAGTGATAACTCTATTTTCTCGTGCAGCTGGACTGCACAA	1739
Qy	1320	TCGGTACACAGAAAAATTCACCTTCGGGGTGTGATCCTGGATATACACTTTTCTCTCACAGGC	1379
Db	1740	TCGATACACAGAAAAATTCACCTTCGGAGGTGATCCTGGATATACATTTTCTCTCAGGC	1799
Qy	1380	TGACTTTCTAGTGTGATCTTTTTTATCCCAAGGCTGTTCGGGTTGCTTTATGAAATCATGCA	1439
Db	1800	AGACTTCCTAGTGTGATCTTTTTTATCCCAAGGCTGTTCGAGTTGCTTTATGAAATTTATGCA	1859
Qy	1440	AACCCCTGCATCCTGCATGCCCTCTCGGAACTCTTCCAACTCTTTGGATGACATCTACTATTTGG	1499
Db	1860	AACACTACATCCTGCATGCCCTCTCGAAACTTCCOATCTTTAGATGACATCTACTATTTTGG	1919
Qy	1500	AGGCCAAAATGCCCAACAATCAGATTTGGTGTTHATTCCTCACAAACCTTCGAAGTGAAGGGA	1559
Db	1920	GGGCCAAGATGCCCAACAATCAAAATTCGCAPTTTATGCTCACCAACCCCGAACTCAGATGA	1979
Qy	1560	AATTTCAATCGAACCTGGAGATATCATTTGGTGTGGCTTGGAAAACATTTGGGATGCTTTATTC	1619
Db	1980	AATTTCCATCGAACCTGGAGATATCATTTGGTGTGGCTTGGAAATCATTTGGGATGGCTTATTC	2039
Qy	1620	TAAAGGTATCAACAGAAAACTTGGAAAAACAGCGCTTATATCCCTCTACAAAGTCCGAGA	1679
Db	2040	TAAAGGTATCAACAGGAAATTTGGGAGGACGGGCTTATATCCCTCTACAAAGTTCGAGA	2099
Qy	1680	GAAGATAGAAACAGTCAAGTATCCCAATATCTCGAAGCTGAAAAATA	1727
Db	2100	GAAGATAGAAACCGTCAAGTATCCCAATATCTCGAGGCTGAAAAATA	2147

```

RESULT 5
US-09-839-136-1
; Sequence 1, Application US/(98)39136
; Patent No. US20020081694A1
; GENERAL INFORMATION:
; APPLICANT: Naoyuki TANIGUCHI et al.
; TITLE OF INVENTION: ALPHA 1-6 FUCOSYLTRANSFERASE
; FILE REFERENCE: 2356-7
; CURRENT APPLICATION NUMBER: US/(09/839,136
; CURRENT FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 09/442,629
; PRIOR FILING DATE: 1999-11-18
; PRIOR APPLICATION NUMBER: 08/913,805
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: PCT/JP97/00171
; PRIOR FILING DATE: 1997-01-23
; NUMBER OF SEQ ID NOS: 15
; SEQUENCE: FastSeq for Windows Version 4.0
; SEQ ID NO 1

```

```

; LENGTH: 1728
; TYPE: DNA
; ORGANISM: Pig
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1728)
US-09-839-136-1

Query Match      84.9%; Score 1467.8; DB 9; Length 1728;
Best Local Similarity 90.6%; Pred No. 0;
Matches 1565; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 1 ATGCGGCGATGACGTGGTTCCTGGCGTGGATTATGCTCATTCCTTTTGGCTGGGGGACC 60
Db 1 ATGCGGCGATGACGTGGTTCCTGGCGTGGATTATGCTCATTCCTTTTGGCTGGGGGACC 60

QY 61 TTGCTATTTTATATAGTGGTTCATTTGGTTCAGATAATGACCCCTGATCACTCCAGC 120
Db 61 TTGCTATTTTATATAGTGGTTCATTTGGTTCAGATAATGACCCCTGATCACTCCAGC 120

QY 121 AGAGAACTCTCAAGATTCTTCAAAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 180
Db 121 CGAGAACTCTCAAGATTCTTCAAAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 180

QY 181 AGGCGAATGGCTGAGTCTCTCCGAATACAGAAAGCCCCCATTTGACCGGGGACAGCTACA 240
Db 181 AGGCGAATGGCTGAGTCTCTCCGAATACAGAAAGCCCCCATTTGATCAGGGGCGAGCTTCA 240

QY 241 GGAAGAGTCCGTGTTTGAAGAAACAGCTTGTTAAGGCCCAAGACAGATTGAAAATTAC 300
Db 241 GGAAGAGTCCGTGTTTGAAGAAACAGCTTGTTAAGGCCCAAGACAGATTGAAAATTAT 300

QY 301 AAGAAACAAAGCTAGAAATGGTCTGGGGAAGGATCATGAATCTTAAAGAGGAGATTGAA 360
Db 301 AAGAAACAAAGCTAGAAATGGTCTGGGGAAGGATCATGAATCTTAAAGAGGAGATTGAA 360

QY 361 AATGGAGCTAAAGAGCTCTGGTCTTTTCTCAAAAGCGAACTGAAGAAATTAAGCATTTA 420
Db 361 AATGGAGCTAAAGAGCTCTGGTCTTTTCTCAAAAGCGAACTGAAGAAATTAAGCATTTA 420

QY 421 GAAGGAATGAATCTCAAAAGACATGAGATGAATTCCTTTGGATTATAGGACACCATGAA 480
Db 421 GAAGGAATGAATCTCAAAAGACATGAGATGAATTCCTTTGGATTATAGGACATCATGAA 480

QY 481 AGGTCTATCATCAGATCTATACCTACCTCAGTCAAAAGAGTGGAGGAGGATGGCGT 540
Db 481 AGGTCTATTAAGCGATCTATACCTACCTCAGTCAAAAGAGTGGAGGAGGATGGCGT 540

QY 541 GAAAAGAGGCGCAAGATCTGACAGAGCTGGTCCAGCGGAGAAATACATATCTCCAGAA 600
Db 541 GAAAAGAGGCGCAAGATCTGACAGAGCTGGTCCAGCGGAGAAATACATATCTCCAGAA 600

QY 601 CTAAGGAGTGCAGAAAGCCAGGAGAGCTGGTGTGTAAATCAATTAAGGCTGGGCTAT 660
Db 601 CCAAGGAGTGCAGAAAGCCAGGAGAGCTGGTGTGTAAATCAATTAAGGCTGGGCTAT 660

QY 661 GTTGTGCAACTCCATCAGCTGCTTACTGTTTCTGATGTTGCTTATGGCAACCCAGCGAACA 720
Db 661 GGTCTGTACGCTCCATCATGTAGTGTACTGCTTTTATGATTTGATATGGCAACCCAGCGAACA 720

QY 721 CTCATCTTTGGAATCTCAGAAATTTGGCGCTATGCTACTGGTGGATGGAGACTGTGTTTGA 780
Db 721 CTCGCTTTGGAATCTCAGAAATTTGGCGCTACGCTACTGGGGGATGGAAACTGTGTTTGA 780

QY 781 CTTGTAAGTGCAGATATGACAGAGATCTGCGCTCTCCATGGGACACTGTTCAGGTGAA 840
Db 781 CTTGTAAGTGCAGATATGACAGAGATCTGCGAGCTCCATGGGACACTGTTCAGGTGAA 840

QY 841 GTAAATGACAAAAACATTCAAGTGGTGCAGCTCCCATTTGTAGACAGCTCCATCCTCGG 900
Db 841 GTAAAGGACAAAATGTTTCAGGTGGTGTGAGCTCCCATTTGTAGACAGGTTCATCTCGT 900

QY 901 CCTCCTTACTTACCAGTGGCTGTTCCAGAAAGACCTTGCAGACCGGACTCCTAAGAGTCCAT 960
```

```

Db 901 CCTCATATTTACCCCTGGCTGTCCAGAAAGACCTTGCAGATCGACTTGTACGAGTCCAT 960
QY 961 GGTGACCTGTCAGTGTGGTGGTGTCCAGTTCCTCAAAATATCTGATTCGTCCCAACCT 1020
Db 961 GGTGATCTGTCAGTGTGGTGGTGTCCAGTTCCTCAAGTTCGTCAAGTACCTGATTCGCCCAACCC 1020
QY 1021 TGGCTGAAAAAGGAAATAGAAAGCCCAAGAAAGCTTGGCTTCAAAATATCTGATTCAT 1080
Db 1021 TGGCTGAAAAAGGAAATAGAAAGCCCAAGAAAGCTTAGGCTTCAAAATATCTGATTCAT 1080
QY 1081 GGAGTTCATGTTCACAGCCACAGACAAAGTGGGAACAGAGCCTTCCACCCATCCAG 1140
Db 1081 GGAGTTCATGTTCAGACCCACAGACAAAGTGGGAAGCGGAGCCTTCCATCCCATGAG 1140
QY 1141 GAGTACATGGTTCACAGCTTGAAGAAATTTTGGCCACAGATGACCCCTGCTTTGTTAAAGAGGCAAAACAAAG 1200
Db 1141 GAATACAGGTGTCAGCTTGAAGAAAGCTTTCAGCTTCTTGGCTCGCAGAAATGCAGTGGAT 1200
QY 1201 AAAAAAGAGATATATCTGGCTACTGATGATCTTACTTTGTTTAAAGAGGCAAAAGCAAAAG 1260
Db 1201 AAAAAAGGCTGTATTTGGCCACAGATGACCCCTGCTTTGTTTAAAGAGGCAAAACAAAG 1260
QY 1261 TACTCCAAATATGAATTTATTTAGTGATACTCTATTTCTTGGTCAGCTGGGACTACACAT 1320
Db 1261 TACCCAGTATGATGAATTTATTTAGTGATACTCTATCTTGGTCAGCTGGGACTACATAT 1320
QY 1321 CGGTACACAGAAAAATTCACCTTCGGGGTGTGATCCTGGATATACACTTTCTCTCACAGGCT 1380
Db 1321 CGATATACAGAAAAATTCACCTTCGGGGTGTGATCCTGGATATACACTTTCTCTCCAGGCA 1380
QY 1381 GACTTCTTAGTGTACTTTTTTCACTCCAGGTCTGTCCGGTGTCTTATGAAATCATGCAA 1440
Db 1381 GACTTCTTAGTGTACTTTTTTCACTCCAGGTCTGTAGAGTGTCTTATGAAATCATGCAA 1440
QY 1441 ACCGTGATCTGATGCTCTGCGAATCTTCATTTCTTTGGATGACATCTACTATTTTGA 1500
Db 1441 GCGTGTGATCTGATGCTCTGCGAATCTTCGGAATCTTCGTTGGATGACATCTACTATTTTGA 1500
QY 1501 GGCCAAATGCCCAATTCAGATTCCTGTTTATCTCACAACCTCGAACTGAAAGAGAA 1560
Db 1501 GGCCAAATGCCCAATTCAGATTCCTGTTTATCTCACAACCTCGAACTGAAAGAGAA 1560
QY 1561 ATTCCAATGGAACTCGAGATATCATTTGGTGTGCTGGAAACCATTTGGATGGTATTCCT 1620
Db 1561 ATCCCATGGAACTCGAGATATCATTTGGTGTGCTGGAAATCACTGGATGGCTATCCT 1620
QY 1621 AAAGGTATCAACAGAAAACTTGGAAAAACAGGCTTATATCCCTCTACAAAGTCCGAGAG 1680
Db 1621 AAAGGTGTAAACAGAAAACTTGGAAAGGACGGGCTTATATCCCTCTACAAAGTTCGAGAG 1680
QY 1681 AAGATAGAAACAGTCAAGATATCCACATATCTCTGAAGCTGAAAAATA 1727
Db 1681 AAGATAGAAACAGTCAAGATATCCACATATCTCTGAAGCTGAAAGTA 1727
```

```

RESULT 6
US-09-971-773-7
; Sequence 7, Application US/09971773
; Publication No. US20030115614A1
; GENERAL INFORMATION:
; APPLICANT: Yutaka KANDA
; APPLICANT: Mitsuo SATOH
; APPLICANT: Kazuyasu NAKAMURA
; APPLICANT: Kazuhisa UCHIDA
; APPLICANT: Toyohide SHINKAWA
; APPLICANT: Naoko YAMANE
; APPLICANT: Motoo YAMASAKI
; APPLICANT: No. US20030115614A1uo HANAI
; TITLE OF INVENTION: ANTIBODY COMPOSITION-PRODUCING CELL
; FILE REFERENCE: 249-202
; CURRENT APPLICATION NUMBER: US/09/971,773
; CURRENT FILING DATE: 2002-08-30
```

;; PRIOR APPLICATION NUMBER: JP 2000-308526  
;; PRIOR FILING DATE: 2000-10-06  
;; PRIOR APPLICATION NUMBER: US 60/268,926  
;; PRIOR FILING DATE: 2001-02-16  
;; NUMBER OF SEQ ID NOS: 73  
;; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 979  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
US-09-971-773-7

Query Match 53.9%; Score 931; DB 11; Length 979;

Best Local Similarity 96.9%; Pred. No. 1.9e-262;

Matches 949; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy	720	ACTCATCTTGAATCTCAGAATTCGAGTGGCGCTATGCTACTGGTGGATGGGAGACTGTTTAG	779
Db	1	ACTCATCTTGAATCTCAGAATTCGAGTGGCGCTATGCTACTGGTGGATGGGAGACTGTTTAG	60
Qy	780	ACCTGTAAGTCAGACATGTACAGACAGATCTGGCTCTCCACTGCACACTGGTCAGGTGA	839
Db	61	ACCTGTAAGTCAGACATGTACAGACAGATCTGGCTCTCCACTGCACACTGGTCAGGTGA	120
Qy	840	AGTAAATGACAAAAACATTCAGTGGTGCAGTCCCAATTTGTAGACAGCTCCATCTCTCG	899
Db	121	AGTAAATGACAAAAACATTCAGTGGTGCAGTCCCAATTTGTAGACAGCTCCATCTCTCG	180
Qy	900	GCCTCTTACTTACCACTGGCTGTTCCAGAGACCTTTCAGACAGCTTCCTAAGAGTCCA	959
Db	181	GCCTCTTACTTACCACTGGCTGTTCCAGAGACCTTTCAGACAGCTTCCTAAGAGTCCA	240
Qy	960	TGGTGACCTCGAGTGTGGTGTCCAGTCTTGTCAAAATCTTGTTCGTCACAAACC	1019
Db	241	TGGTGACCTCGAGTGTGGTGTCCAGTCTTGTCAAAATCTTGTTCGTCACAAACC	300
Qy	1020	TTGGCTGAAAAGGAAATAGAGAGCCACCAAGAGCTTGGCTTCAAAATCTCAGTTAT	1079
Db	301	TTGGCTGAAAAGGAAATAGAGAGCCACCAAGAGCTTGGCTTCAAAATCTCAGTTAT	360
Qy	1080	TGGAGTCCATGTACAGCGCACAGACAAAGTGGGAAACAGAGCAGCTTCCACCCATCGA	1139
Db	361	TGGAGTCCATGTACAGCGCACAGACAAAGTGGGAAACAGAGCAGCTTCCACCCATCGA	420
Qy	1140	GGAGTACATGTACAGTGTGAAGACATTTTCAGTCTTCGACGACGAAATGCAAGTGA	1199
Db	421	AGAGTACATGTACAGTGTGAAGACATTTTCAGTCTTCGACGACGAAATGCAAGTGA	480
Qy	1200	TAAAAAAGAGTATATCTGGCTACTGTATGATCCTCTTGTAAAGGAGGCAAAAGACAAA	1259
Db	481	TAAAAAAGAGTATATCTGGCTACTGTATGATCCTCTTGTAAAGGAGGCAAAAGACAAA	540
Qy	1260	GTACTCCAAATATGAATTTATAGTGAATCTCTATTTCTTGGTCAGTGAATACAAA	1319
Db	541	GTACTCCAAATATGAATTTATAGTGAATCTCTATTTCTTGGTCAGTGAATACAAA	600
Qy	1320	TCGGTACACAGAAATTCACCTCGGGGTGTATCTGTGATATACATTTCTCTCAGGC	1379
Db	601	TCGGTACACAGAAATTCACCTCGGGGTGTATCTGTGATATACATTTCTCTCAGGC	660
Qy	1380	TGACTTTCTAGTGTGATCTTTTTCATCCAGGTCTGTGGGTGCTTATGAAATCATGCA	1439
Db	661	TGACTTTCTAGTGTGATCTTTTTCATCCAGGTCTGTGGGTGCTTATGAAATCATGCA	720
Qy	1440	AACCTCGCATCTGTATGCTCTCGGAATCTCCATTTCTTGGATGACATCTACTATTTGG	1499
Db	721	AACCTCGCATCTGTATGCTCTCGGAATCTCCATTTCTTGGATGACATCTACTATTTGG	780
Qy	1500	AGGCCAAAATGCCCAATCAGATGTCTTTATCTCACAACCTCGAATGAGAGGA	1559
Db	781	AGGCCAAAATGCCCAATCAGATGTCTTTATCTCACAACCTCGAATGAGAGGA	840
Qy	1560	AATTCCAATGGAACCTGGAGATATCATTTGGTGTGGCTGCGAAACCATTTGGGATGTTATTC	1619

Db	841	AATTCCAATGGAACCTGGAGATATCATTTGGTGTGGCTGGAACCATTTGGGATGTTATTC	900
Qy	1620	TAAAGGTATCAACAGAAAACTTGGAAAAACAGGCTTATATCCCTCTACAAAGTCCGAGA	1679
Db	901	TAAAGGTATCAACAGAAAACTTGGAAAAACAGGCTTATATCCCTCTACAAAGTCCGAGA	960
Qy	1680	GAAGATAGAAACAGTCAAG	1698
Db	961	GAAGATAGAAACGCTCAAG	979

# RESULT 7

US-09-971-773-6  
; Sequence 6, Application US/09971773  
; Publication No. US20030115614A1  
; GENERAL INFORMATION:  
; APPLICANT: Yutaka KANDA  
; APPLICANT: Mitsuo SATOH  
; APPLICANT: Kazuyasu NAKAMURA  
; APPLICANT: Kazuhisa UCHIDA  
; APPLICANT: Toyohide SHINKAWA  
; APPLICANT: Naoko YAMANE  
; APPLICANT: Motoo YAMASAKI  
; APPLICANT: No. US20030115614A1uo HANAI  
; TITLE OF INVENTION: ANTIBODY COMPOSITION-PRODUCING CELL  
; FILE REFERENCE: 249-202  
; CURRENT APPLICATION NUMBER: US/09/971,773  
; CURRENT FILING DATE: 2002-08-30  
; PRIOR APPLICATION NUMBER: JP 2000-308526  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 60/268,926  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 979  
; TYPE: DNA  
; ORGANISM: Cricetulus griseus  
US-09-971-773-6

Query Match 51.7%; Score 894.2; DB 11; Length 979;

Best Local Similarity 94.6%; Pred. No. 1.2e-251;

Matches 926; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy	720	ACTCATCTTGAATCTCAGAATTCGCGCTATGCTACTGGTGGATGGGAGACTGTTTAG	779
Db	1	ACTCATCTTGAATCTCAGAATTCGCGCTATGCTACTGGGAGATGGGAGACTGTTTAG	60
Qy	780	ACCTGTAAGTCAGACATGTACAGACAGATCTGGCTCTCCACTGCACACTGGTCAGGTGA	839
Db	61	ACCTGTAAGTCAGACATGTACAGACAGATCTGGCTCTCCACTGCACACTGGTCAGGTGA	120
Qy	840	AGTAAATGACAAAAACATTCAGTGGTGCAGTCCCAATTTGTAGACAGCTCCATCTCTCG	899
Db	121	AGTAAATGACAAAAACATTCAGTGGTGCAGTCCCAATTTGTAGACAGCTCCATCTCTCG	180
Qy	900	GCCTCTTACTTACCACTGGCTGTTCCAGAGACCTTTCAGACAGCTTCCTAAGAGTCCA	959
Db	181	TCCTCTTACTTACCACTGGCTGTTCCAGAGACCTTTCAGAGACCTTCCTGAGAGTCCA	240
Qy	960	TGGTGACCTCGAGTGTGGTGTCCAGTCTTGTCAAAATCTTGTTCGTCACAAACC	1019
Db	241	TGGTGACCTCGAGTGTGGTGTCCAGTCTTGTCAAAATCTTGTTCGTCACAAACC	300
Qy	1020	TTGGCTGAAAAGGAAATAGAGAGCCACCAAGAGCTTGGCTTCAAAATCTCAGTTAT	1079
Db	301	TTGGCTGAAAAGGAAATAGAGAGCCACCAAGAGCTTGGCTTCAAAATCTCAGTTAT	360
Qy	1080	TGGAGTCCATGTACAGCGCACAGACAAAGTGGGAAACAGAGCAGCTTCCACCCATCGA	1139
Db	361	TGGAGTCCATGTACAGCGCACAGACAAAGTGGGAAACAGAGCAGCTTCCACCCATCGA	420



```
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 13359
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO ALL09847.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
US-09-864-761-13359

Query Match 15.6%; Score 269.2; DB 9; Length 503;
Best Local Similarity 89.8%; Pred. No. 3.4e-68;
Matches 289; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 1406 CCCAGTCTGTGGGTTGTTATGAAATCATGCAAAACCTGATCCTGATGCTCTGCGA 1465
Db 6 CCAAGTCTGTGGGTTGTTATGAAATCATGCAAAACCTGATCCTGATGCTCTGCGA 65

QY 1466 ACTTCATCTTTGATGACATCTACTATTTTGGAGGCCAAATGCCACAAATCAGATTG 1525
Db 66 ACTTCATCTTTGATGACATCTACTATTTTGGAGGCCAAATGCCACAAATCAGATTG 125

QY 1526 CTGTTATCTCTCACAACTGCAACTGAAGAGGAATTCCAATGGAACCTGGAGATCA 1585
Db 126 CCAATTTATGCTTACCACCCGCACTGAGATGAAATTTCCCATGGAACCTGGAGATCA 185

QY 1586 TTGGTGTGGCTGGAACCACTGGGATGTTATCTAAAGGTATCAACAGAAAACCTTGAA 1645
Db 186 TTGGTGTGGCTGGAACCACTGGGATGTTATCTAAAGGTATCAACAGAAAACCTTGAA 245

QY 1646 AAACAGGCTTATATCCCTCTCAAAAGTCCGAGAGAGATGAAACAGTCAAGTATCCCA 1705
Db 246 GGAAGGCTTATATCCCTCTCAAAAGTCCGAGAGAGATGAAACAGTCAAGTATCCCA 305

QY 1706 CATATCTGAGCTGAAAATA 1727
Db 306 CATATCTGAGCTGAAAATA 327

RESULT 10
US-09-864-761-13292
; Sequence 13292, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aomics-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
```

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 13292
; LENGTH: 551
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO ALL09847.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
US-09-864-761-13292

Query Match 12.2%; Score 210.8; DB 9; Length 551;
Best Local Similarity 80.0%; Pred. No. 5.2e-51;
Matches 248; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 773 TGTTAGACCTGTAAAGTGAGACATGTACAGACAGATCTGGCTCTCCACTGGACACTGT 832
Db 116 TGTGAATGGTGGATGTATAGGAATACCATGTGTATATATATATATATATATAT 175

QY 833 CAGGTGAAGTAAATGACAAAACATTCAGTGGTGGAGCTCCCATTTAGACAGCTCC 892
Db 176 CAGGTGAAGTGAAGGACAAAATTTCAAGTGGTGGAGCTCCCATTTAGACAGCTCT 235

QY 893 ATCTCGGCTCTTACTTACTTACCTGCTGTTCCAGAGAGCTTCAGACCGACTCTAA 952
Db 236 ATCCCGGCTCTTACTTACTTACCTGCTGTTCCAGAGAGCTTCAGACCGACTCTAC 295

QY 953 GAGTCCATGTTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1012
Db 296 GAGTCCATGTTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 355

QY 1013 CACAACTTGGCTGGAAGAAATAGAAAGCCACCAAGAGCTTGGCTTCAAAATC 1072
Db 356 CACAGCTTGGCTGGAAGAAATAGAAAGCCACCAAGAGCTTGGCTTCAAAATC 415

QY 1073 CAGTATTGG 1082
Db 416 CAGTATTGG 425

RESULT 11
US-09-864-761-18181
; Sequence 18181, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aomics-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
```

FILE REFERENCE: Aeonica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 18181  
LENGTH: 248  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO ALL09847.1  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.2  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.6  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.1  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.9  
OTHER INFORMATION: SWISSPROT HIT: Q9VLZ7, EVALUE 6.50e+00  
OTHER INFORMATION: NT HIT: AF038280.1, EVALUE 0.00e+00  
OTHER INFORMATION: EST\_HUMAN HIT: AW387766.1, EVALUE 0.00e+00  
US-09-864-761-18181

Query Match 11.9%; Score 205.4; DB 9; Length 248;  
Best Local Similarity 89.5%; Pred. No. 1.2e-49;  
Matches 221; Conservative 0; Mismatches 26; Indels 0; Gaps 0;  
QY 836 GTGAAGTAATGACAAAACATTCAGTGGTCCAGTCCCACTTGTAGACAGCCTCCATC 895  
DB 1 GTGAAGTGAAGGACAAAATGTTCAAGTGGTCCAGTCCCACTTGTAGACAGTCTTCATC 60  
QY 896 CTCGGCTCTTACTTACCACTGGCTGTCAGAGACCTTCACAGACCTCCCTAAG 955  
DB 61 CCCGCTCCCTATTTACCCCTTGGCTGTACCAAGAGACCTCGCAGATCGACTGTGACG 120

QY 956 TCCATGTCACCTCCAGTGTGGTGGTGTCCAGTTGTCAAAATCTTGTATTCGTCCAC 1015  
DB 121 TGCATGTCACCTCCAGTGTGGTGGTGTCTCAGTTGTCAAAATCTTGTATTCGTCCAC 180  
QY 1016 AACCTTGGCTGAAAAGAAATAGAAGAACCCACCAAGAGCTTGGCTTCAAAATCCAG 1075  
DB 181 AGCTTGGCTAGAAAAGAAATAGAAGAACCCACCAAGAGCTTGGCTTCAAAATCCAG 240  
QY 1076 TTATTGG 1082  
DB 241 TTATTGG 247

## RESULT 12

US-09-864-761-30284  
Sequence 30284, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
FILE REFERENCE: Aeonica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 30284  
LENGTH: 248  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO ALL09847.3  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4



OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3  
OTHER INFORMATION: SWISSPROT HIT: Q9VLZ7, EVALUE 6.50e+00  
OTHER INFORMATION: EST HUMAN HIT: AW387766.1, EVALUE 0.00e+00  
OTHER INFORMATION: NT HIT: AF038280.1, EVALUE 0.00e+00  
US-09-864-761-30284

Query Match 11.9%; Score 205.4; DB 9; Length 248;  
Best Local Similarity 89.5%; Pred. No. 1.2e-49;  
Matches 221; Conservative 0; Mismatches 26; Indels 0; Gaps 0;  
QY 836 GTGAAGTAATGACAAACAAATTCAAGTGGTCGAGCTCCCATTTGTAGACAGCTCCATC 895  
DB 1 GTGAAGTGAAGGACAAACAAATGTTCAAGTGGTCGAGCTCCCATTTGTAGACAGCTCTCATC 60  
QY 896 CTCGCGCTCTTACTTACCACCTGGCTGTTCCAGAGACCTTGCAGACGACTCTCCTAAGAG 955  
DB 61 CCGTCTCCCATATTTACCTTGGCTGTACAGAGACCTCGCAGATCGATTTGTACGAG 120  
QY 956 TCCATGTGACCTCGACGTGTGGTGGTGTCCCAAGTTTGTCAAACTATTGATTCGTCAC 1015  
DB 121 TGCATGTGTGACCTCGACGTGTGGTGGTGTCTCAGTTTGTCAAACTATTGATTCGTCAC 180  
QY 1016 AACCTTGCTGGAAGAAAGAAATAGAGAGACCAAGACCTTGGCTTCAAACTATCCAG 1075  
DB 181 AGCCTTGCTAGAAAAGAAATAGAGAGACCAAGACCTTGGCTTCAAACTATCCAG 240  
QY 1076 TTATTGG 1082  
DB 241 TTATTGG 247

## RESULT 13

US-09-971-773-3  
; Sequence 3, Application US/09971773  
; Publication No. US200301156141  
; GENERAL INFORMATION:  
; APPLICANT: Yutaka KANDA  
; APPLICANT: Mitsuo SATOH  
; APPLICANT: Kazuyasu NAKAMURA  
; APPLICANT: Kazuhisa UCHIDA  
; APPLICANT: Toyohide SHINKAWA  
; APPLICANT: Naoko YAMANE  
; APPLICANT: Motoo YAMASAKI  
; APPLICANT: No. US200301156141uo HANAI  
; TITLE OF INVENTION: ANTIBODY COMPOSITION-PRODUCING CELL  
; FILE REFERENCE: 249-202  
; CURRENT APPLICATION NUMBER: US/09/971,773  
; CURRENT FILING DATE: 2002-08-30  
; PRIOR APPLICATION NUMBER: JP 2000-308526  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 60/268,926  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 9196  
; TYPE: DNA  
; ORGANISM: Cricetulus griseus  
US-09-971-773-3

Query Match 10.9%; Score 188.6; DB 11; Length 9196;  
Best Local Similarity 95.6%; Pred. No. 1e-43;  
Matches 194; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 1 ATCGGGGATGAGTGGTTCCTGGCGTTGGATTATGCTCATTTCTTTTTCCTGGGGACC 60  
DB 2549 ATCGGGGATGAGTGGTTCCTGGCGTTGGATTATGCTCATTTCTTTTTCCTGGGGACC 2608  
QY 61 TTGTTATTTTATATAGTGGTTCATTTGGTTCGAGATATGACCCCTGATCATCTCCAGC 120  
DB 2609 TTATTGTTTATATAGTGGTTCATTTGGTTCGAGATATGACCCCTGATCATCTTAGC 2668  
QY 121 AGAGAAGTCTCAAGATTCTTGCAAGCTTGAAACGCTTTAAACAGCAAAATGAAGACTTG 180

DB 2669 AGAGAAGTCTCCAGATTCTTGCAAGCTGCGCTTTAAACAGCAAAATGAAGACTTG 2728  
QY 181 AGCGGAATGGCTGAGTCTCTCCG 203  
DB 2729 AGGGAATGGCTGAGTCTCTCCG 2751

## RESULT 14

US-10-125-968-274  
; Sequence 274, Application US/10125968  
; Publication No. US20030215805A1  
; GENERAL INFORMATION:  
; APPLICANT: Lillie, James  
; APPLICANT: Palermo, Adam  
; APPLICANT: Wang, Youzhen  
; APPLICANT: Steinmann, Kathleen  
; APPLICANT: Elias, Josh  
; APPLICANT: Mertens, Maureen  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND  
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF BREAST CANCER  
; FILE REFERENCE: MRI-032  
; CURRENT APPLICATION NUMBER: US/10/125,968  
; CURRENT FILING DATE: 2002-04-19  
; PRIOR APPLICATION NUMBER: US 60/285,163  
; PRIOR FILING DATE: 2001-04-20  
; NUMBER OF SEQ ID NOS: 1417  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 274  
; LENGTH: 481  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 204, 266, 320, 328, 402, 424, 462, 470  
; OTHER INFORMATION: n = A,T,C or G  
US-10-125-968-274

Query Match 8.5%; Score 147.4; DB 13; Length 481;  
Best Local Similarity 84.5%; Pred. No. 2e-32;  
Matches 212; Conservative 0; Mismatches 33; Indels 6; Gaps 4;  
QY 239 CAGGAAGAGTCCGCTGTTTGAAGAACAGCAGCTTGTAAAGGCCAAAGAACAG-ATTGAAAAT 297  
DB 19 CGGCGAGGTACGCGTTTACAAAGCAGCAGCTTGTAAAGGCCAAAGAACAGTATTGAAAAT 78  
QY 298 TACAAGAAACAG--CTAGAAATGCTCTGGGAAAGGATCATGAAATCTTAAGAGGAGGA 355  
DB 79 TACAAGAAACAGACACAGTAAATGCTCTGGGAAAGGATCATGAAATCTCTGAGGAGGAGGA 138  
QY 356 TTGAAATGGAGCTTAAAGAGCTCTGGTTTTTCTACAAGCGCACTGAAGAAATTAAGC 415  
DB 139 TTGAAATGGAGCTTAAAGAGCTCTGGTTTTTCTACAAGAGTGAATTAAGAAATTAAGA 198  
QY 416 ATTTAGAAGGAATGAATGAACTCCAAAGACATGC-AGATGAAATCTTTTGG--ATTAGGAC 472  
DB 199 ACTTAAAGGAATGAATGAACTCCAAAGACATGCAAGATGATTTCTTTGGGATTTAGGAC 258  
QY 473 ACCATGAAGG 483  
DB 259 ATCATGANAAG 269

## RESULT 15

US-09-864-761-1421  
; Sequence 1421, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng

;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
;; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

;; FILE REFERENCE: A6mics-X-1

;; CURRENT APPLICATION NUMBER: US/09/864,761

;; CURRENT FILING DATE: 2001-05-23

;; PRIOR APPLICATION NUMBER: US 60/180,312

;; PRIOR FILING DATE: 2000-02-04

;; PRIOR APPLICATION NUMBER: US 60/207,456

;; PRIOR FILING DATE: 2000-05-26

;; PRIOR APPLICATION NUMBER: US 09/632,366

;; PRIOR FILING DATE: 2000-08-03

;; PRIOR APPLICATION NUMBER: GB 24263.6

;; PRIOR FILING DATE: 2000-10-04

;; PRIOR APPLICATION NUMBER: US 60/236,359

;; PRIOR FILING DATE: 2000-09-27

;; PRIOR APPLICATION NUMBER: PCT/US01/00666

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00667

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00664

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00669

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00665

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00668

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00663

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00662

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00661

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00670

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: US 60/234,687

;; PRIOR FILING DATE: 2000-09-21

;; PRIOR APPLICATION NUMBER: US 09/608,408

;; PRIOR FILING DATE: 2000-06-30

;; PRIOR APPLICATION NUMBER: US 09/774,203

;; PRIOR FILING DATE: 2001-01-29

;; NUMBER OF SEQ ID NOS: 49117

;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

;; SEQ ID NO 1421

;; LENGTH: 384

;; TYPE: DNA

;; ORGANISM: Homo sapiens

;; FEATURE:

;; OTHER INFORMATION: MAP TO AL109847.1

;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.2

;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5

;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6

;; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 3.6

;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5

;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.1

;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3

;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7

;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3

;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.9

;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.9

;; US-09-864-761-1421

Query Match 7.3%; Score 126.8; DB 9; Length 384;

Best Local Similarity 73.9%; Pred. No. 26-26;

Matches 161; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy 773 TCTTTAGACCTTAAGTAGACATGTACAGACAGATCTGGCTCTCCACTGGACACTGGT 832

Db 167 TCTTGAATGGTGGATGTATAGGAATACCATGTGGTAAATGATTATGTTCATATATTGT 226

Qy 833 CAGGTGAAGTAATGACAAACATTCAGTGGTGGCTCCCATTTGTAGACAGCTCC 892

Db 227 CAGGTGAAGTAAGGACAAATGTTCAAGTGGTGGCTCCCATTTGTAGACAGCTCTC 286

Qy 893 ATCCTCGCCTCCTTACTTACCACCTGGCTGTTCCAGAGACCTTGCAGACCGACTCTAA 952

Db 287 ATCCCCGCTCTCCATATTTACCCCTTGGCTGTACCAAGACCTCGCAGATCGACTTGTAC 346

Qy 953 GAGTCCATGGTGACCCCTGCGAGTGTGGTGGGTGTCCCGAG 990

Db 347 GAGTGCATGGTGACCCCTGCGAGTGTGGTGGGTGTCTCAG 384

Search completed: February 2, 2004, 19:03:05

Job time : 616.522 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: February 2, 2004, 11:54:20 ; Search time 3711.31 Seconds  
(without alignments)  
11316.242 Million cell updates/sec

Title: US-09-971-773-2  
Perfect score: 1728  
Sequence: 1 atggggcatgactgttc.....atctgaagctgaaataag 1728

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0  
Searched: 22781392 seqs, 12152238056 residues  
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: em\_estba.\*

2: em\_esthum.\*

3: em\_estin.\*

4: em\_estmu.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_htc.\*

9: gb\_est1.\*

10: gb\_est2.\*

11: gb\_htc.\*

12: gb\_est3.\*

13: gb\_est4.\*

14: gb\_est5.\*

15: em\_estfun.\*

16: em\_estom.\*

17: em\_gss\_hum.\*

18: em\_gss\_inv.\*

19: em\_gss\_pln.\*

20: em\_gss\_vrt.\*

21: em\_gss\_fun.\*

22: em\_gss\_mam.\*

23: em\_gss\_mus.\*

24: em\_gss\_pro.\*

25: em\_gss\_rod.\*

26: em\_gss\_phg.\*

27: em\_gss\_vrl.\*

28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1728	100.0	2766	11 AK051811	AK051811 Mus muscu
2	1728	100.0	3052	11 AK048520	AK048520 Mus muscu
3	913.2	52.8	937	13 BQ892101	BQ892101 AGENCOURT
4	801.6	46.4	898	14 CB194828	CB194828 AGENCOURT

5	787.6	45.6	1066	12	BM928521	BM928521 AGENCOURT
6	747.2	43.2	940	13	BQ653329	BQ653329 AGENCOURT
7	746.8	43.2	900	13	BM928521	BM928521 AGENCOURT
8	735.2	42.5	886	13	BQ423860	BQ423860 AGENCOURT
9	725	42.0	923	13	BQ423921	BQ423921 AGENCOURT
10	720.4	41.7	855	13	BQ184296	BQ184296 AGENCOURT
11	704.6	40.8	937	13	BQ513162	BQ513162 AGENCOURT
12	696.8	40.3	1027	13	BQ513162	BQ513162 AGENCOURT
13	687.4	39.8	834	9	AU124128	AU124128 AGENCOURT
14	660.2	38.2	702	12	BQ973221	BQ973221 AGENCOURT
15	656	38.0	656	13	BQ552954	BQ552954 AGENCOURT
16	627.2	36.3	752	13	BQ604588	BQ604588 AGENCOURT
17	627.2	36.3	752	13	BQ604588	BQ604588 AGENCOURT
18	620.2	35.9	965	12	B1546364	B1546364 AGENCOURT
19	615.2	35.6	790	14	CB989769	CB989769 AGENCOURT
20	613	35.5	613	14	CA539475	CA539475 AGENCOURT
21	608.8	35.2	692	14	CB105611	CB105611 AGENCOURT
22	608.8	35.2	692	14	CB105611	CB105611 AGENCOURT
23	593	34.3	805	14	CA489662	CA489662 AGENCOURT
24	592.8	34.3	717	14	CB322023	CB322023 AGENCOURT
25	576.4	33.4	769	10	BG501191	BG501191 AGENCOURT
26	568.2	32.9	573	9	AI122135	AI122135 AGENCOURT
27	567.8	32.9	936	10	BE893107	BE893107 AGENCOURT
28	562.8	32.6	646	14	CB105920	CB105920 AGENCOURT
29	559	32.3	715	12	BG771085	BG771085 AGENCOURT
30	557.2	32.2	900	13	BQ731540	BQ731540 AGENCOURT
31	554.2	32.1	938	13	BQ433236	BQ433236 AGENCOURT
32	551	31.9	970	13	BQ926824	BQ926824 AGENCOURT
33	550.6	31.9	637	14	CB105617	CB105617 AGENCOURT
34	548.2	31.7	731	13	BQ456265	BQ456265 AGENCOURT
35	542	31.4	622	14	CB105602	CB105602 AGENCOURT
36	541	31.3	621	14	CB105873	CB105873 AGENCOURT
37	538	31.1	618	14	CB105618	CB105618 AGENCOURT
38	535.2	31.0	653	10	BG619839	BG619839 AGENCOURT
39	531.8	30.8	607	14	CB105608	CB105608 AGENCOURT
40	531.8	30.8	607	14	CB105882	CB105882 AGENCOURT
41	531.8	30.8	608	14	CB105918	CB105918 AGENCOURT
42	530.2	30.7	607	14	CB105899	CB105899 AGENCOURT
43	529.8	30.7	534	10	BE690071	BE690071 AGENCOURT
44	524.2	30.3	610	14	CB105867	CB105867 AGENCOURT
45	512.4	29.7	586	14	CB105886	CB105886 AGENCOURT

ALIGNMENTS

RESULT 1

AK051811

LOCUS

DEFINITION

AK051811

2766 bp

mus musculus 12 days embryo eyeball cDNA, RIKEN full-length

enriched library, clone:D330003K17 product: fucosyltransferase 8,

full insert sequence.

AK051811

AK051811.1

GI:26342229

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

99279253

10349636

REFERENCE

2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itch, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

20499374

PUBMED

REFERENCE  
AUTHORS

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861

REFERENCE  
AUTHORS

4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,  
Arakawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,  
Saito, K., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,  
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,  
Kuehl, P., Lewis, S., Mateu, Y., Nikaide, I., Pesole, G.,  
Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M.,  
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,  
Bardarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,  
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,  
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,  
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,  
Marchionni, L., Mashima, J., Mazzaletti, J., Mombaerts, P., Nordone, P.,  
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,  
Toyo-oka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L.,  
Wyshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S.  
and Hayashizaki, Y.

TITLE  
JOURNAL

Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)  
21085660  
11217851

REFERENCE  
AUTHORS

5 The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 2766)

TITLE  
JOURNAL

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanaoka, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitho, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)

TITLE  
JOURNAL

CDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/  
Location/Qualifiers

## FEATURES

source 1. 2766  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM,DB:D230003K17"  
/db\_xref="taxon:10090"  
/clones="D230003K17"  
/tissue\_type="eyeball"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="12 days embryo"  
473\_-2200  
/note="unnamed protein product; fucosyltransferase 8  
(MGD|MG1:1858901, GB|NM\_016893, evidence: BLASTN, 99%,  
match=2175)  
putative"  
/codon\_start=1  
/protein\_id="BAC34777.1"  
/db\_xref="GI:26342230"  
/translation="MRATGSGRWIMLIFAWGTLFPYIGHLVRDNDHPDHSRELS  
KILAKRLKQNEEDLRMAESRIPEGIQDGTATGRVLEQLVKAQEIENIKK  
QARNGLKQHEILRRRIENGAKELWFFLOSELKKLEGNELQRLHAEILDLGHHE  
RSIMDLVYLSQTDGAGWREKAKDLTELQRRITVYLPQNKDCSKARKLVNKGK  
GYGCOLHVYVCFMIAVGTORTLILESONRWYATGWTVPVSECTDRLSGSTGH  
WSSEVNDKNQVVELPIVDSLHPPPLPLAVPEDLADRLLVHGDPAVWVSOFVKY  
LIRPQWLEKEIBEATKUGFGKGVIGVHVRITDKVGTAAAFHPIEYMWVERHFQL  
LARMQVKKRVYLATDDPTLLKEAKTKSYNEFISDSWSAGLHNRHNTENSURVG  
ILDHFLSQADFLVCTFSQVRYVEIMQTLHPDASANFHSLLDDIYFGGQNAHQI  
AVVPHKRTREEIPMEPCDIIGVAGNHWGVSKGINRKLKGTGLYPSYKVREKIETVK  
YPTYPEAK"  
polyA\_signal 2750. 2755  
/note="putative"  
2766  
polyA\_site  
/note="putative"  
BASE COUNT 838 a 591 c 619 g 718 t  
ORIGIN  
Query Match 100.0%; Score 1728; DB 11; Length 2766;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1728; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGCGGCGATGGACTGGTCTCTGGCGTTGGATTATGCTCATTTCTTTTTCCTGGGGACC 60  
DB 473 ATGCGGCGATGGACTGGTCTCTGGCGTTGGATTATGCTCATTTCTTTTTCCTGGGGACC 532  
QY 61 TTGTTATTTATATAGTGTGTCATTGTTGTCGAGATAATGACCCCTGATCCTCCAGC 120  
DB 533 TTGTTATTTATATAGTGTGTCATTGTTGTCGAGATAATGACCCCTGATCCTCCAGC 592  
QY 121 AGAGAACTCTCCAGATTTCTTGCAAGCTTGAACCGCTTAAACAGCAAAATGAAGACTTG 180  
DB 593 AGAGAACTCTCCAGATTTCTTGCAAGCTTGAACCGCTTAAACAGCAAAATGAAGACTTG 652  
QY 181 AGCGAATGGCTGAGTCTCTCCGAATACCAAGCGCCCATTTGACCGGGGACAGCTACA 240  
DB 653 AGCGAATGGCTGAGTCTCTCCGAATACCAAGCGCCCATTTGACCGGGGACAGCTACA 712  
QY 241 GGAAGAGTCCCTGTTTGTAGAGAAACAGCTTCTTAAAGCCAAAGAACAGATTGAAATTTAC 300  
DB 713 GGAAGAGTCCCTGTTTGTAGAGAAACAGCTTCTTAAAGCCAAAGAACAGATTGAAATTTAC 772  
QY 301 AAGAAAACAGCTAGAAATGGTCTGGGAAGGATCATGAAATCTTAAAGAGAGAGATTGAA 360  
DB 773 AAGAAAACAGCTAGAAATGGTCTGGGAAGGATCATGAAATCTTAAAGAGAGAGATTGAA 832  
QY 361 AATGAGCTAAGAGCTCTGCTTTTCTCAAGAGCGAAGCTGAAGAAATTTAAAGCATTTA 420  
DB 833 AATGAGCTAAGAGCTCTGCTTTTCTCAAGAGCGAAGCTGAAGAAATTTAAAGCATTTA 892  
QY 421 GAAGAAATGAATCAATCCCAAGACATGCAGATGAAATCTTTTGGATTTAGGACACCATGAA 480  
DB 893 GAAGAAATGAATCAATCCCAAGACATGCAGATGAAATCTTTTGGATTTAGGACACCATGAA 952  
QY 481 AGGCTATCATGACAGATCTTACTACTCTAGTCAACAGATGAGGACAGGGGATTTGGCGT 540

[illegible]

Db 2033 ATTCAATGGAACCTGGAGATATCATTTGGTGTGGCTGGAAACCATTTGGGATGGTTATCTT 2099

QY 1621 AAAGGTATCAACAGAGAAAACCTTGGAAAAACAGAGCTTATATCCCTCTACAAAGTCCGAGAG 1680  
|||||  
Db 2093 AAAGGTATCAACAGAGAAAACCTTGGAAAAACAGGCTTATATCCCTCTACAAAGTCCGAGAG 2152  
|||||

QY 1681 AGATAGAAACAGTCAAGTATCCACATATCTTGAAGCTGAAAATAG 1728  
|||||  
Db 2153 AGATAGAAACAGTCAAGTATCCACATATCTTGAAGCTGAAAATAG 2200  
|||||

RESULT 2

AK048520

LOCUS

DEFINITION Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched library, clone:CI30068P22 product:fucoyltransferase 8, full insert sequence.

ACCESSION AK048520

VERSION AK048520.1

KEYWORDS GI:26339371

SOURCE HTC; CAP trapper.

ORGANISM Mus musculus (house mouse)

REFERENCE 1

AUTHORS Carninci, P. and Hayashizaki, Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 9279253

PUBMED 10349636

REFERENCE 2

AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE 20495374

PUBMED 11042159

REFERENCE 3

AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

TITLE RIKEN integrated sequence analysis (RISA) system - 384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL Genome Res. 10 (11), 1757-1771 (2000)

MEDLINE 20530913

PUBMED 11076861

REFERENCE 4

AUTHORS Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Flatschmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Resole, G., Quackenbush, J., Schriml, L. M., Staab, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Batsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodrigues, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohtsuki, S. and Hayashizaki, Y.

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL Nature 409 (6821), 685-690 (2001)

MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

21085660  
11217851

5

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

6 (bases 1 to 3052)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, I., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kohji, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitho, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/  
Location/Qualifiers  
1. .3052  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM\_DB:Cl30068P22"  
/db\_xref="taxon:10090"  
/clone="Cl30068P22"  
/tissue type="head"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="16 days embryo"  
654. .2381  
/note="unnamed protein product; fucosyltransferase 8 (MGD|MGI:1858901, GB|NM\_016893, evidence: BLASTN, 99%, match=2175)  
putative"

CDS

/codon\_start=1  
/protein\_id="BAC33357.1"  
/db\_xref="GI:26339372"  
/translation="MRWNTGSRWIMLILFAWGTLLFYIGHLVRDNDPHDSRELS  
KILAKLERLKOQNEDLREMAESLRIPGPIDQGTATGVRVLEQLVKAEQIENYKK  
QARNGLDKHEILRRRIENGAKELWFLQSELKKLHLEGNELORHADEILLDGHHE  
RSIMTDLTYLSDTDGADWRKEAKDLTELVRRIITLQNPDCSKARKVCNINKEG  
GYGQQLHVYQVITAYGQRTLILESQNRWYATGGWETVFRPSETCDRSLSYGH  
WSGEYNDKNIQVLELP1VDSLHPPYPLFLAVPEDLADRLRHVGDPAVMVVSQFVKY  
LIRPOWLEKIEEATKGLPKFKHPIGVHVRRTKVGTEAAPHPIEYMWVHEHFOL  
LARMQVKKVYLATDPTLLKKAATKYSNVEPTSDNSISWSAGLHNRVYENSILRGV  
ILDIHLSQADFLVCTFSQVCRVAYEIMQTLHPDASANFSLDDIYYPFGQNAHQI  
AVYPKPRTEIEIPMPGDIIGVAGNHWGYSKGNIRKLGKIGLYPSPYKVKREKLETVK  
YPTYPEAK"

polyA\_signal  
3036. .3041  
/note="putative"

polyA\_site  
3052  
/note="putative"

BASE COUNT 895 a 685 c 684 g 788 t

ORIGIN

Query Match 100.0%; Score 1728; DB 11; Length 3052;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1728; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCGGGCAGTGGACTGGTCTCTGGCGTTGGATTATGCTCATTCTTTTGGCTGGGGGACC 60  
Db 654 ATGCGGGCAGTGGACTGGTCTCTGGCGTTGGATTATGCTCATTCTTTTGGCTGGGGGACC 713  
Qy 61 TTGTTATTTTATATAGTGGTTCATTTGCTTCGAGATAATGACCCCTGTGATCCTCCAGC 120  
Db 714 TTGTTATTTTATATAGTGGTTCATTTGCTTCGAGATAATGACCCCTGTGATCCTCCAGC 773  
Qy 121 AGAGAACTCTCCAAAGATTTCTTGGAAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 180  
Db 774 AGAGAACTCTCCAAAGATTTCTTGGAAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 833  
Qy 181 AGGCGAATGGCTGAGTCTCTCCGAATACCAAGAGCCCATTTGACCGGGGACAGCTACA 240  
Db 834 AGGCGAATGGCTGAGTCTCTCCGAATACCAAGAGCCCATTTGACCGGGGACAGCTACA 893  
Qy 241 GGAAGAGTCCGTGTTTGTAGAGAAACAGCTTCTTAAGGCCAAAGAACAGATTGAAAATTTAC 300  
Db 894 GGAAGAGTCCGTGTTTGTAGAGAAACAGCTTCTTAAGGCCAAAGAACAGATTGAAAATTTAC 953  
Qy 301 AAGAAACCAAGCTAGAAATGGTCTCGGGAAGGATCATGAAATCTTTAAGAGGAGGATTGAA 360  
Db 954 AAGAAACCAAGCTAGAAATGGTCTCGGGAAGGATCATGAAATCTTTAAGAGGAGGATTGAA 1013  
Qy 361 AATGAGCTAAGAGCTCTGGTTTTTCTCAAGCCGAACTGAAGAAATTAAGCATTTA 420  
Db 1014 AATGAGCTAAGAGCTCTGGTTTTTCTCAAGCCGAACTGAAGAAATTAAGCATTTA 1073  
Qy 421 GAAGCAATGAATCCAAAGACATGCAGATGAAATCTTTTGGATTTTAGGACACCATGAA 480  
Db 1074 GAAGCAATGAATCCAAAGACATGCAGATGAAATCTTTTGGATTTTAGGACACCATGAA 1133  
Qy 481 AGGTCTATCATGACAGATCTATCTACTCAGTCAAAACAGATGAGCAGGAGGATTGGCGT 540  
Db 1134 AGGTCTATCATGACAGATCTATCTACTCAGTCAAAACAGATGAGCAGGAGGATTGGCGT 1193  
Qy 541 GAAAGAGGCCAAGATCTGACAGAGCTGGTCAGCGGAGAAATAACATATCTCCAGAT 600  
Db 1194 GAAAGAGGCCAAGATCTGACAGAGCTGGTCAGCGGAGAAATAACATATCTCCAGAT 1253  
Qy 601 CCTAAGGACTGCACAAAGCCAGGAAGCTGGTGTGTAAACATCAATAAAGGCTGTGGCTAT 660  
Db 1254 CCTAAGGACTGCACAAAGCCAGGAAGCTGGTGTGTAAACATCAATAAAGGCTGTGGCTAT 1313  
Qy 661 GGTGTCAACTCCATCACTGCTGCTACTGTTTCATGATTTGTCACCCAGGAGAAACA 720  
Db 1314 GGTGTCAACTCCATCACTGCTGCTACTGTTTCATGATTTGTCACCCAGGAGAAACA 1373  
Qy 721 CTCACTTGGATCTCAGAAATTTGGCGCTATGCTACTGTTGGATGGAGACTGTGTTTGA 780  
Db 1374 CTCACTTGGATCTCAGAAATTTGGCGCTATGCTACTGTTGGATGGAGACTGTGTTTGA 1433  
Qy 781 CCTGTAAGTGAGACATGTACAGACAGATCTGGCCCTCTCCACTGGACACTGCTCAGGTGAA 840  
Db 1434 CCTGTAAGTGAGACATGTACAGACAGATCTGGCCCTCTCCACTGGACACTGCTCAGGTGAA 1493  
Qy 841 GTAAATGACAAAAACATTCAGAGTGGTCCAGCTCCCCATTTGTAGACAGCTCCATCCTCGG 900  
Db 1494 GTAAATGACAAAAACATTCAGAGTGGTCCAGCTCCCCATTTGTAGACAGCTCCATCCTCGG 1553  
Qy 901 CCTCTTACTTACCACTGGCTGTTCCAGAGACCTTGCAGACCCGACTCCTTAAGAGTCCAT 960  
Db 1554 CCTCTTACTTACCACTGGCTGTTCCAGAGACCTTGCAGACCCGACTCCTTAAGAGTCCAT 1613  
Qy 961 GGTGACCTCGAGTGTGGTGGGTGTCAGGTTTGTCAAATAATCTTGAATCGTCCCAACCT 1020  
Db 1614 GGTGACCTCGAGTGTGGTGGGTGTCAGGTTTGTCAAATAATCTTGAATCGTCCCAACCT 1673

QY 1021 TGGCTGGAAGGAATAGAGAGCCACCAGAGCTTGGCTTCAAAATCCAGTTATT 1080  
Db 1674 TGGCTGGAAGGAATAGAGAGCCACCAGAGCTTGGCTTCAAAATCCAGTTATT 1733  
QY 1081 GGAGTCCATGTCAGACGACAGACAAAGTGGACAGAGCAGCTTCCACCCATCGAG 1140  
Db 1734 GGAGTCCATGTCAGACGACAGACAAAGTGGACAGAGCAGCTTCCACCCATCGAG 1793  
QY 1141 GAGTACATGGTACAGCTTGAAGACATTTTCAGCTTTCGCGACGAGAAATGCAAGTGGAT 1200  
Db 1794 GAGTACATGGTACAGCTTGAAGACATTTTCAGCTTTCGCGACGAGAAATGCAAGTGGAT 1853  
QY 1201 AAAAAAGAGTATATCTGGCTACTGAGATCCTACTTTTAAAGAGGCGCAAGCAAG 1260  
Db 1854 AAAAAAGAGTATATCTGGCTACTGAGATCCTACTTTTAAAGAGGCGCAAGCAAG 1913  
QY 1261 TACTCCAATTATGAATTTATGATGATAAATCTATTTCTTGGTCAGCTGGATACACAT 1320  
Db 1914 TACTCCAATTATGAATTTATGATGATAAATCTATTTCTTGGTCAGCTGGATACACAT 1973  
QY 1321 CGGTACACAGAAATTCACCTCGGGGTGTGATCTCGGATATACACTTTCTCTCAGAGCT 1380  
Db 1974 CGGTACACAGAAATTCACCTCGGGGTGTGATCTCGGATATACACTTTCTCTCAGAGCT 2033  
QY 1381 GACTTTCTAGTGTGATCTTTTTCATCCAGAGTCTGTGGGTGTGCTTATGAATCATGCAA 1440  
Db 2034 GACTTTCTAGTGTGATCTTTTTCATCCAGAGTCTGTGGGTGTGCTTATGAATCATGCAA 2093  
QY 1441 ACCCTGCAATCTGATGCTCTGCGAATCTCAATCTTTTGGATGATCACTATTTTGGGA 1500  
Db 2094 ACCCTGCAATCTGATGCTCTGCGAATCTCAATCTTTTGGATGATCACTATTTTGGGA 2153  
QY 1501 GGCCAAATGCCCAATCAGATTGCTGTTTATCTCTCAAACTCGAACTGGAAGAGAA 1560  
Db 2154 GGCCAAATGCCCAATCAGATTGCTGTTTATCTCTCAAACTCGAACTGGAAGAGAA 2213  
QY 1561 ATTCCAATGGAACTCGAGATATCAATGCTGTGGTGGAAACCAATGGGATGGTTATTCT 1620  
Db 2214 ATTCCAATGGAACTCGAGATATCAATGCTGTGGTGGAAACCAATGGGATGGTTATTCT 2273  
QY 1621 AAGGTATCAACAGAAAACTTGGAAAAACAGGCTTATATCCCTCAAAATCGGAGAG 1680  
Db 2274 AAGGTATCAACAGAAAACTTGGAAAAACAGGCTTATATCCCTCAAAATCGGAGAG 2333  
QY 1681 AAGTAGAAGACGTCAAGTATCCCATATCTCTGAGCTGGAATAG 1728  
Db 2334 AAGTAGAAGACGTCAAGTATCCCATATCTCTGAGCTGGAATAG 2381

RESULT 3  
LOCUS BQ892101  
DEFINITION BQ892101 NIH\_MGC\_129 Mus musculus cDNA clone IMAGE:6389725  
5', mRNA sequence.  
ACCESSION BQ892101  
VERSION BQ892101.1 GI:22284115  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
1 (bases 1 to 937)  
NIH-MGC <http://mgc.mci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)  
Tissue Procurement: Susan L. Sullivan, PhD.  
cDNA Library Preparation: ResGen, Invitrogen Corp  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILLNL at:

<http://image.llnl.gov>  
Plate: L1AM13875 row: d column: 14  
High quality sequence start: 130  
High quality sequence stop: 780.  
FEATURES  
Location/Qualifiers  
source  
1..937  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6389725"  
/lab\_host="NIH\_MGC\_129"  
/clone\_id="NIH\_MGC\_129"  
/note="Organ: olfactory epithelium; Vector:  
pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2: NotI; Cloned  
unidirectionally. Primer: Oligo dt. Average insert size  
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this  
is a NIH MGC Library." 244 t 1 others  
BASE COUNT 268 a 215 c 209 g 244 t  
ORIGIN  
Query Match 52.8%; Score 913.2; DB 13; Length 937;  
Best Local Similarity 99.5%; Pred. No. 4.5e-226;  
Matches 926; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
QY 736 CAGAAATGGCGCTATGCTACTGGTGGATGGGAGACTGTGTTAGACCTGTAAAGTGAACA 795  
Db 1 CAGAAATGGCGCTATGCTACTGGTGGATGGGAGACTGTGTTAGACCTGTAAAGTGAACA 60  
QY 796 TGATACAGACAGATCTGCGCTCTCCACTGGACACTGGTCAAGTGAATTAATGACAAAAAC 855  
Db 61 TGATACAGACAGATCTGCGCTCTCCACTGGACACTGGTCAAGTGAATTAATGACAAAAAC 120  
QY 856 ATTCAAGTGGTGGAGTCCGCCATTTGTAGACAGCTCCATCTCGGCCCTCTTACTTACCA 915  
Db 121 ATTCAAGTGGTGGAGTCCGCCATTTGTAGACAGCTCCATCTCGGCCCTCTTACTTACCA 180  
QY 916 CTGCTGCTTCAGAAACCTTTGCAGACCGACTCTCTAAGAGTCCATGAGTGCACCTGCAGTG 975  
Db 181 CTGCTGCTTCAGAAACCTTTGCAGACCGACTCTCTAAGAGTCCATGAGTGCACCTGCAGTG 240  
QY 976 TGGTGGGTGCTCCAGTTGTCAAAATACCTTGTCTGTCACAACTTGGCTGGAAAAAGAA 1035  
Db 241 TGGTGGGTGCTCCAGTTGTCAAAATACCTTGTCTGTCACAACTTGGCTGGAAAAAGAA 300  
QY 1036 ATAGAAGAGCCACCAAGAGCTTGGCTTCAAAATCCAGTTATTGGAGTCCATGTGACA 1095  
Db 301 ATAGAAGAGCCACCAAGAGCTTGGCTTCAAAATCCAGTTATTGGAGTCCATGTGACA 360  
QY 1096 CGCACAGACAAAGTGAGAACAGAGCAGCTTCCACCCCATCGAGGAGTACATGTTACAC 1155  
Db 361 CGCACAGACAAAGTGAGAACAGAGCAGCTTCCACCCCATCGAGGAGTACATGTTACAC 420  
QY 1156 GTTCAAGAACATTTTACGCTTCTCGACGAGAGTCAAGTGGATATAAAGAGAGTATAT 1215  
Db 421 GTTCAAGAACATTTTACGCTTCTCGACGAGAGTCAAGTGGATATAAAGAGAGTATAT 480  
QY 1216 CTGCTACTGATGATCTACTTTGTTAAGAGGAGCAAGAGCAAAAGTACTCCAATATGAA 1275  
Db 481 CTGCTACTGATGATCTACTTTGTTAAGAGGAGCAAGAGCAAAAGTACTCCAATATGAA 540  
QY 1276 TTTATTAGTGATACTCTATTCTTTGGTCACTGAGCTACACAAATCGGTACACAGAAAT 1335  
Db 541 TTTATTAGTGATACTCTATTCTTTGGTCACTGAGCTACACAAATCGGTACACAGAAAT 600  
QY 1336 TCACCTCGGGGTGTGATCTGGATATACATTTCTCTCAGAGGTGACTTTCTAGTGTGT 1395  
Db 601 TCACCTCGGGGTGTGATCTGGATATACATTTCTCTCAGAGGTGACTTTCTAGTGTGT 660  
QY 1396 ACTTTTTCATCCAGAGTCTGTGCGGTGCTTATGAATCATGCAACCTGCATCCTGAT 1455  
Db 661 ACTTTTTCATCCAGAGTCTGTGCGGTGCTTATGAATCATGCAACCTGCATCCTGAT 720  
QY 1456 GCCTCTCGAACTTCTCTTTCTTGGATGACATCTACTATTTTGGAGGCCAAATGCCCCAC 1515



Db 721 GGCCTCGGAACCTTCATCTTTGGATGACATCTACTATTTTGGAGGCCAAATGCCAC 780  
QY 1516 AATCAGATTGCTGTTTATCTCACAACCTCGAACTGAAGAGAAATTCGAATGAACT 1575  
Db 781 AATCAGATTGCTGTTTATCTCACAACCTCGAACTGAAGAGAAATTCGAATGAACT 840  
QY 1576 GGAGATATCATTTGGTGTGGCTGGAAC-ATTGGATGTTTATCTAAAGGTATCAACAG 1634  
Db 841 GGAGATATCATTTGGTGTGGCTGGAACCAATGGGATGTTTATCTAAAGGTATCAACAG 900  
QY 1635 AAAAAGTGAAGAAACAGGCTTATCCCTCC 1665  
Db 901 AAAAAGTGAAGAAACAGGCTTATCCCTCC 931

RESULT 4  
CB194828  
LOCUS  
DEFINITION  
AGENCOURT\_11259261 NIH\_MGC\_135 Mus musculus cDNA clone  
IMAGE:30136221 5', mRNA sequence.  
ACCESSION  
CB194828  
VERSION  
CB194828.1 GI:28220043  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 898)  
NIH-MGC <http://mgi.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Dr. David Rowe  
cDNA library preparation: Invitrogen Corp  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: NDAM0035 row: h column: 22  
High quality sequence stop: 658.  
Location/Qualifiers  
1..898  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30136221"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 135"  
/note="Vector: pCMVSPORT6.1; Site 1: EcoRV; Site 2: NotI;  
Normalized full-length enriched library from pooled mouse  
embryonic limb, maxilla and mandible, day 12.5, 13.5, 14.5  
, and 15.5 (size selected for the 0.5-1 kb fragments)  
Cloned directionally, priming method: Oligo-dT. cDNA  
enrichment: >1k bp. Average insert size 1.6k bp.  
Normalization (Cot value): 7.5 kb. Priming sequence:  
5'-GACTAGTTCTAGATCGGAGCGCGCCGCTTT-3' Tissue contributed by  
, David Rowe. Library constructed by Resgen, Invitrogen  
Corp."

## FEATURES

source  
BASE COUNT 272 a 203 c 193 g 230 t  
ORIGIN

Query Match 46.4%; Score 801.6; DB 14; Length 898;  
Best Local Similarity 98.8%; Pred. No. 4.6e-197;  
Matches 818; Conservative 0; Mismatches 9; Indels 1; Gaps 1;  
QY 899 GGCCTCCTTACTTACCACTGGCTGTTCCAGAGACCTTGACAGCGACTCTCAAGAGTCC 958  
Db 1 GGCCTCCTTACTTACCACTGGCTGTTCCAGAGACCTTGACAGCGACTCTTAAAGTCC 60  
QY 959 ATGGTGACCCCTGAGTGGTGGGTGCCAGTTTGTCAATACATTGATCGTCCACAC 1018

Db 61 ATGTGACCCCTGCGAGTGTGGGGTGTCCTCCAGTTTGTCAATATCTGATTCGTCCACAC 120  
QY 1019 CTTGGCTGGAAAAGGAAATAGAGAAGCCCAAGAAGCTTTGGCTTCAACATCCAGTTA 1078  
Db 121 CTTGGCTGGAAAAGGAAATAGAGAAGCCCAAGAAGCTTTGGCTTCAACATCCAGTTA 180  
QY 1079 TTGGAGTCCATGTGAGACGCAAGCAAAAGTGGGAACAGAAGAGCCCTTCCACCCCATCG 1138  
Db 181 TTGGAGTCCATGTGAGACGCAAGCAAAAGTGGGAACAGAAGAGCCCTTCCACCCCATCG 240  
QY 1139 AGGAGTACATGGTACACGTTTGAAGAACAATTTTACGCTTCTCGCAGCGAGAATGCAAGTGG 1198  
Db 241 AGGAGTACATGGTACACGTTTGAAGAACAATTTTACGCTTCTCGCAGCGAGAATGCAAGTGG 300  
QY 1199 ATAAAAAAGAGTATATCTGGCTACTGATGATCCCTACTTTGTTAAAGAGGCAAGACAA 1258  
Db 301 ATAAAAAAGAGTATATCTGGCTACTGATGATCCCTACTTTGTTAAAGAGGCAAGACAA 360  
QY 1259 AGTACTCCAAATTATGAATTTATTAGTGATTAATCTATTTCTTTGGTCACTGAGCTACACA 1318  
Db 361 AGTACTCCAAATTATGAATTTATTAGTGATTAATCTATTTCTTTGGTCACTGAGCTACACA 420  
QY 1319 ATCGGTACACAGAAATTCACCTTCGGGGTGTGATCCCTGGATATACACTTTCTCTCACAGG 1378  
Db 421 ATCGGTACACAGAAATTCACCTTCGGGGTGTGATCCCTGGATATACACTTTCTCTCACAGG 480  
QY 1379 CTGACTTTCTAGTGTGACTTTTTTATCCCAAGGCTCTGTGGGGTGTGCTTATGAATCATGC 1438  
Db 481 CTGACTTTCTAGTGTGACTTTTTTATCCCAAGGCTCTGTGGGGTGTGCTTATGAATCATGC 540  
QY 1439 AAACCCCTGCATCTGATGCTCTCGAACTTCCATTTCTTTGGATGACATCTACTATTTTG 1498  
Db 541 AAACCCCTGCATCTGATGCTCTCGAACTTCCATTTCTTTGGATGACATCTACTATTTTG 600  
QY 1499 GAGGCCAAATGCCCAATCAGATTCGTTTATCTCTCACAACCTCGAACTGGAAGAGG 1558  
Db 601 GAGGCCAAATGCCCAATCAGATTCGTTTATCTCTCACAACCTCGAACTGGAAGAGG 660  
QY 1559 AAATTCCAATGGAACTCGAGATATCATTTGGTGTGGCTGGAAACCATTTGGATGTTATT 1618  
Db 661 AAATTCCAATGGAACTCGAGATATCATTTGGTGTGGCTGGAAACCATTTGGATGTTATT 720  
QY 1619 CTAAAGGTATCAACAGAAAACCTTGGAACCAAGGCTTATATCCCTCTTCAAAAGTCCAG 1678  
Db 721 CTAAAGGTATCAACAGAAAACCTTGGAACCAAGGCTTATATCCCTCTTCAAAAGTCCAG 780  
QY 1679 AGAAGTAGAAGACAG-TCAAGTATCCCATATCTCGAAGCTGAAAAA 1725  
Db 781 AGAAGTAGAAGACCACTTCAAGTATCCCATATCTCGAAGCTGAAAAA 828

RESULT 5  
BM928521  
LOCUS  
DEFINITION  
AGENCOURT\_6715254 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:5798262  
5', mRNA sequence.  
ACCESSION  
BM928521  
VERSION  
BM928521.1 GI:19378900  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 1066)  
NIH-MGC <http://mgi.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: CGAP (Stanford)  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1CM2025 row: p column: 07  
 High quality sequence stop: 666.

Location/Qualifiers

1. .1066  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5798262"  
 /tissue\_type="hepatocellular carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 100"  
 /note="Organ: liver; Vector: pOTB7; Site: 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGCACGAG(G). Size-selected >500bp for average insert size  
 1.8kb. Library constructed by Ling Hong in the laboratory  
 of Gerald M. Rubin (University of California, Berkeley)  
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
 II RT (Life Technologies). Note: this is a NIH\_MGC  
 Library."

BASE COUNT 315 a 235 c 248 g 264 t 4 others

ORIGIN

Query Match 45.6%; Score 787.6; DB 12; Length 1066;

Best Local Similarity 89.2%; Pred. No. 2.1e-193; Mismatches 103; Indels 3; Gaps 3;

Matches 880; Conservative 0;

QY 366 AGCTAAAGAGCTCTGGTCTTTTCTCAAGCGCACTGAAGAAATTAAGCACTTTAGAAGG 425

DB 1 AGCTAAAGAGCTCTGGTCTTTTCTCAAGCGCACTGAAGAAATTAAGCACTTTAGAAGG 60

QY 426 AAATGAATCCAAAGACATGCGAGATGAATTTCTTTGGATTGAGACACCATGAAGGTC 485

DB 61 AAATGAATCCAAAGACATGCGAGATGAATTTCTTTGGATTGAGACATCATGAAGGTC 120

QY 486 TATCATGACAGATCTATCTACTCTCAGTCAACACATGAGCGGGATTGGCGTGAAGA 545

DB 121 TATAATGACGGATCTATCTACTCTCAGTCAACACATGAGCGGGATTGGCGGGAAGA 180

QY 546 AGAGGCCAAAGATCTGACAGAGCTGGTCCAGCGGAGATAACATATCTCCAGAACTCTAA 605

DB 181 AGAGGCCAAAGATCTGACAGAGCTGGTCCAGCGGAGATAACATATCTCCAGAACTCTAA 240

QY 606 GGACTGCGAGCAAGCCAGGAAGCTGGTGTAAACATCAATAAAGCTGTGGCTATGGTTG 665

DB 241 GGACTGCGAGCAAGCCAGGAAGCTGGTGTAAACATCAATAAAGCTGTGGCTATGGTTG 300

QY 666 TCAATCCATCAGTGGTCTACTGTTTCTATGATTCGTTATGACCCAGCGACACTCAT 725

DB 301 TCAGCTCCATCATGTGGTCTACTGTTTCTATGATTCGTTATGACCCAGCGACACTCAT 360

QY 726 CTTGGAATCTCAGATTGGCGTCTATGTTGATGGGAGACTGTTTACACCTGT 785

DB 361 CTTGGAATCTCAGATTGGCGTCTATGTTGATGGGAGACTGTTTACACCTGT 420

QY 786 AAGTGAGACATGTACAGACAGATCTGGCTCTCCACTGGACACTGGTCAAGTGAATAA 845

DB 421 AAGTGAGACATGTACAGACAGATCTGGCTCTCCACTGGACACTGGTCAAGTGAATAA 480

QY 846 TGACAAAACATTAAGTGGTTCGAGCTCCCATTTGATGACAGCTCCATCTCGGCTCC 905

DB 481 GGACAAAACATTAAGTGGTTCGAGCTCCCATTTGATGACAGCTCCATCTCGGCTCC 540

QY 906 TTACTTACCACTGGCTGTTCAGAGAGACTTCGACAGCTCCCTAAGAGTCCATGGGA 965

DB 541 ATATTACCTTGGCTGTACCAAGAGACTTCGACAGCTCCCTAAGAGTCCATGGGA 600

QY 966 CCCTGCAGTGGTGGGTGTCCTCAGTTGTTCAATACTTGTTCGTCACAACTTGGCT 1025

|||||

DB 601 CCCTGCAGTGGTGGGTGCTCTCAGTTGTCAAAATACATCTGATCCGCCACAGCCTTGGCT 660

QY 1026 GGAAAGGAAATAGAAAGAGCCACCACAAAGCTTGGCTTTCAAAACATCCAGTTATTGGAGT 1085

DB 661 AGAAAGGAAATAGAAAGAGCCACCACAAAGCTTGGCTTTCAAAACATCCAGTTATTGGAGT 720

QY 1086 CCATGTGACGCGACAGACAAAGTGGGAACAGAGCAGCCTTCCACCCATCGAGAGTA 1145

DB 721 CCATGTGACGCGACAGACAAAGTGGGAACAGAGCAGCCTTCCATCCATCCATTTGAAGAGTA 780

QY 1146 CATGGTACACCTTGAAGAACATTTTTCAGCTTCTCGCAGCAGAGATGCAAGTGGATATAAAA 1205

DB 781 CATGTGTCATGTTGAAGAACATTTTTCAGCTTCTCGCAGCAGAGATGCAAGTGGATATAAAA 840

QY 1206 AAGAGTATATCTGGTCTACTGATGATCTCTTCTTTTAAAGGAGGCAAAAGCAAACTACTC 1265

DB 841 AAGAGTATATCTGGTCTACTGATGATCTCTTCTTTTAAAGGAGGCAAAAGCAAACTACTC 900

QY 1266 CAATTATGAATTT-ATTAGTCAATCTATTT-CTTGGTCACTGGGACTACAACTCGG 1323

DB 901 CAATTATGAATTTAATAATGATAACTCTATTTCTCTGTCAGCGGAGCTTGCCTCATCGN 960

QY 1324 -TACACAGAAATTCACCTTCGGGGTG 1348

DB 961 ATACCCCGAAAAATTCCTTCGGGG 986

RESULT 6

BQ653329

LOCUS

DEFINITION BQ653329 940 bp mRNA linear EST 15-JUL-2002  
 AGENCOURT 8298428 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:6270458  
 5', mRNA sequence.

ACCESSION

VERSION BQ653329.1 GI:21777501

KEYWORDS

SOURCE EST.

ORGANISM Homo sapiens (human)

Homology

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 940)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: CGAP (Stanford)

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L1CM2445 row: k column: 03

High quality sequence stop: 700.

Location/Qualifiers

1. .940

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:6270458"

/tissue\_type="hepatocellular carcinoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_100"

/note="Organ: liver; Vector: pOTB7; Site: 1: XhoI; Site 2:

EcoRI; cDNA made by oligo-dT priming. Directionally cloned

into EcoRI/XhoI sites using the following 5' adaptor:

GGCACGAG(G). Size-selected >500bp for average insert size

1.8kb. Library constructed by Ling Hong in the laboratory

of Gerald M. Rubin (University of California, Berkeley)

using ZAP-cDNA synthesis kit (Stratagene) and Superscript

II RT (Life Technologies). Note: this is a NIH\_MGC

Library."

BASE COUNT 286 a 197 c 222 g 235 t



```
Db 364 TGTCAATACTTGTATCCGCCACACGCTTGCTAGAAAAAGAAATAGAAAGACCAACAA 423
Qy GAAGCTTGGCTTCAAAATCAGTATTTGGAGTCCATGTCAGACGACAGACAAAGTGGG 1112
Db |||||||
Qy 424 GAAGCTTGGCTTCAAAATCAGTATTTGGAGTCCATGTCAGACGACAGACAAAGTGGG 483
Db |||||||
Qy 1113 AACAGAGCAGCTTCCACCCCATCGAGGAGTACATGGTACACGTTGAAGAACATTTTCA 1172
Db |||||||
Qy 484 AACAGAGCTGCTTCCATCCATTTGAAGAGTACATGGTGCATGTTGAAGAACATTTTCA 543
Db |||||||
Qy 1173 GCTTCTCGCAGCAGAAATGCAAGTGGATAAAAAAGAGTATATCTGGCTACTGATGCC 1232
Db |||||||
Qy 544 GCTTCTCGCAGCAGAAATGCAAGTGGACAAAAAAGAGTATTTGGCCACAGATGACCC 603
Db |||||||
Qy 1233 TACTTTTGTAAAGAGGAGGCAAGACAAAGTACTTCCAAATTTATGATGATACTC 1292
Db |||||||
Qy 604 TCTTTTAAAGAGGAGGCAAAACAAAGTACCCCAATTTATGATGATACTC 663
Db |||||||
Qy 1293 TATTTCTTGGTCACTGGACTACCAATCGGTACACAGAAAAATTCATTCGGGGTGTGAT 1352
Db |||||||
Qy 1353 CCTGATATACATTTCTCTCACAGGCTGACTTTCTAGTGTGATCTTTTCTATCCAGGT 1412
Db |||||||
Qy 724 CCTGATATACATTTCTCTCTCAAGCAGACTTTCTAGTGTGATCTTTTCTATNCCCAGT 783
Db |||||||
Qy 1413 CTGTCGGGTGCTTATGAATCATGCAAAACCTGCTGATGCTCTGCGAACTTCCA 1472
Db |||||||
Qy 784 CTGGGAGTGTCTTATGAATTTATGCAAACTACATCTGATGCTCTGCAAACTTCCA 843
Db |||||||
Qy 1473 TTTCTTGGATGACATCTACTATTTTGGAGGCCAAATCCCAACATCAGATTGC 1526
Db |||||||
Qy 844 TTCCTTAGATGACATTTACTATTTTGGGGCCAGAAATGCCACAATCAAAATTCG 897
Db |||||||
```

```
RESULT 8
BO423860
LOCUS AGENCOURT_7894821 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6158200
DEFINITION 5', mRNA sequence.
ACCESSION BO423860
VERSION BQ423860.1 GI:21119175
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM13505 row: e column: 17
High quality sequence stop: 641.
FEATURES
location/Qualifiers
1..886
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6158200"
/cisue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 72"
/note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
```

```
Query Match 42.5%; Score 735.2; DB 13; Length 886;
Best Local Similarity 90.7%; Pred. No. 8.4e-180;
Matches 794; Conservative 0; Mismatches 79; Indels 2; Gaps 1;
BASE COUNT 273 a 185 c 212 g 215 t 1 others
ORIGIN
408 ATTAAGACATTTAGAGGAATGAATCCAAAGACATGACAGATCAAAATCTCTTTGGATTT 467
Db 1 ATTAAGACATTTAGAGGAATGAATCCAAAGACATGACAGATCAAAATCTCTTTGGATTT 60
Qy 468 AGGACACCATGAAAGTCTTATCATGACAGATCTATATCTACTTACTCTAGTCAACAGATGGAGC 527
Db 61 AGGACATCATGAAAGTCTTATATGACGGATCTATATCTACTCTAGTCAACAGATGGAGC 120
Qy 528 AGGGATTTGGCTGAAAAGAGGCAAGATCTGACAGAGTGTCTCCAGCGGAGAAATAC 587
Db 121 AGGTGATTGGCGGAAAAGAGGCAAGATCTGACAGAACTGGTTCAGCGGAGAAATAC 180
Qy 588 ATATCTCCAGAACTCTTAAGGACTGACAGAAAGCCAGGAAGCTGTGTGTAACATCAATAA 647
Db 181 ATATCTTCAGAACTCTTAAGGACTGACAGAAAGCCAGGAAGCTGTGTGTAATATCAACA 240
Qy 648 AGGCTGTGGCTATGGTGTCAACTCCATCAAGTGTCTACTTGTTCATGATTGCTTATGG 707
Db 241 AGGCTGTGGCTATGGTGTCAAGTCCATCATGTGTCTACTTGTTCATGATTGCTATGG 300
Qy 708 CACCAGGAAACACTATCTTTGGAATCTCAGAAATGGGGCTATGCTACTTGTGATGGGA 767
Db 301 CACCAGGAAACACTATCTTTGGAATCTCAGAAATGGGGCTATGCTACTTGTGATGGGA 360
Qy 768 GACTGTCTTTAGACCTGTAAGTGAGACATGTACAGACAGATCTGCGCTCTCCACTGGACA 827
Db 361 GACTGTATTTAGGCTTGTAGTGAGACATGTACAGACAGATCTGCGCTCTCCACTGGACA 420
Qy 828 CTGCTCAGGTGAAGTCAATGACAAAAACATTTCAAGTGTGTCAGCTCCCATTTGTAGACAG 887
Db 421 CTGCTCAGGTGAAGTCAAGACAAAAATGTTCAAGTGTGTCAGCTCCCATTTGTAGACAG 480
Qy 888 CCTCCATCTCTGGCTCTCTTACCTTACCACTGGCTGTTCAGAAAGACCTTGCAGACCGACT 947
Db 481 TCTTCATCCCGCTCTCTCATATTTTACCTTGGCTGTACACAGACACCTTCGAGATCGACT 540
Qy 948 CTTAAGAGTCCATGCTGACCTGCACTGTGTGGTGTGTCAGTTCCTCCAGTTTGTCAATCTTCAT 1007
Db 541 TGTACGAGTGCATGGTGACCTTCAGTGTGGTGTGGTGTGTCAGTGTGGTGTGTCAGTGTG 600
Qy 1008 TCGTCCACAACTTTGGCTGGAAAAAGAAATAGAAAGACCCACCAAGAGCTTGGCTTCAA 1067
Db 601 CCGCCACAGCTTGGCTAGAAAAAGAAATAGAAAGAGCCACCAAGAGCTTGGCTTCAA 660
Qy 1068 ACATCCAGTTATTTGGAGTCCATGTGACAGCAGACAGAAAGTGGGAACAGAGAGCTTT 1127
Db 661 ACATCCAGTTATTTGGAGTCCATGTGACAGCAGACAGAAAGTGGGAACAGAGAGCTTT 720
Qy 1128 CCACCCCATCGAGAGTACATGTTACAGTTTGAAGACATTTTTCAGTTCCTCGCAGCAG 1187
Db 721 CCATCCCATTTGAAGAGTACATGTTGTCATGTTTGAAGACATTTTTCAGTTCCTCGCAG 780
Qy 1188 AATGCAAGTGGATAAAAAAGAGTATATCTGGCTACTGATGATCC--TACTTTTGTAAAG 1245
Db 781 AATGCAAGTGGACAAAAAGAGTGTATTTGGCCACAGATGACCCCTCTCTTTATTAAGG 840
Qy 1246 GAGGCAAGACAAAGTACTTCCAAATTTATCAATTTAT 1280
Db 841 GAGGCAAGACAAAGTACTTCCAAATTTATCAATTTT 875
RESULT 9
BX429921
LOCUS BX429921 923 bp mRNA linear EST 15-MAY-2003
```

DEFINITION BX429921 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA  
clone CS0DG005Y007 5-PRIME, mRNA sequence.  
ACCESSION BX429921  
VERSION BX429921.1 GI:30780877  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 923)  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 7499.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0BAA012ZA08 CS01114 1&cluster=7499.f.  
Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID : CS0BAA012ZA08\_CS01114\_1.  
Location/Qualifiers  
1..923  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DG005Y007"  
/tissue\_type="B CELLS (RAMOS CELL LINE)"  
/cell\_line="RAMOS CELL LINE"  
/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"  
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
with a NotI-oligo(dT) primer. Five prime end enriched  
double-strand cDNA was digested with Not I and cloned into  
the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
Library was not normalized."  
BASE COUNT 285 a 184 c 231 g 218 t 5 others  
ORIGIN  
Query Match 42.0%; Score 725; DB 13; Length 923;  
Best Local Similarity 91.3%; Pred. No. 3.8e-177;  
Matches 811; Conservative 0; Mismatches 73; Indels 4; Gaps 4;  
QY 324 GGGGAAGGATCATG-AAATCTTTAAGAGGAGGATTGAAATGGAGCTTAAGAGCTCTGGT 382  
DB 31 GGGNAAGNATCATGNAATCTCTGAGGAGGAGATTGAAATGGAGCTTAAGAGCTCTGGT 90  
QY 383 TTTTCTCAAAACGCACTGGAAGAAATTAAGCATTTTGAAGGAATGAATCCAAAGAC 442  
DB 91 TTTTCTCAAGAGTGAATTTGAAGAAATTAAGAACTTAGAAGAAATGAATCCAAAGAC 150  
QY 443 ATGCAGATGAATTTCTTTTGGATTAGGACACCATGAAGGTCTATCATGACAGATCTAT 502  
DB 151 ATGCAGATGAATTTCTTTTGGATTAGGACATCATGAAAGGTCTATAATGACGGATCTAT 210  
QY 503 ACTACTCTAGTCAACACAGATGAGAGGGGATTGGCGTGAAAGAGCCCAAGATCTGA 562  
DB 211 ACTACTCTAGTCAAGATGAGAGGGGATTGGCGTGAAAGAGCCCAAGATCTGA 270  
QY 563 CAGAGCTGTCCAGGGGGAATTAACATATCTCCAGAATCTTAAGGACTGCAGCAAGCCA 622  
DB 271 CAGAAGCTGTTCAGCGGAGATAACATATCTTCAGAATCCCAAGGACTGCAGCAAGCCA 330  
QY 623 GGAAGCTGTGTGAACATCAATTAAGGCTGTGGCTATGGTGTCACTCCATCATGCTGG 682  
DB 331 AAAAGCTGTGTGTAATATCAACAAAGGCTGTGGCTATGGCTGTGCTCCATCATGTGG 390  
QY 683 TCTACTGTTTCATGATTGCTTATGCCACCCAGCAACTCATCTTGAATCTCAGAAATT 742  
DB 391 TCTACTGTTTCATGATTGCTTATGCCACCCAGCAACTCATCTTGAATCTCAGAAATT 450

QY 743 GGCCTATGCTACTGTTGGATGGGAGACTGTGTTTAGACCTGTAAAGTGAACATGTACAG 802  
DB 451 GGCCTATGCTACTGTTGGATGGGAGACTGTGTTTAGGCCTGTAAAGTGAACATGTACAG 510  
QY 803 ACAGATCTGGGCTCTCCACTGGACACTGGTCCAGGTGAAGTAAATGACAAAAACATTTCAAG 862  
DB 511 ACAGATCTGGGATCTCCACTGGACACTGGTCCAGGTGAAGTGAAGCAAAAAATTTCAAG 570  
QY 863 TGGTCGAGCTCCCATTTGTAGACAGCCTCCATCTCCGCTCTCTTACTTACACATGGCTG 922  
DB 571 TGGTCGAGCTCTCCCATTTGTAGACAGCTCTCATCCCGCTCTCCATATTTACCTTGGCTG 630  
QY 923 TTCAGAAAGACCTTGCAGACCGACTCTTAAGAGTCCATGGTGACCTGCAGTGTGGTGG 982  
DB 631 TACCAGAAAGACCTGCAGATCGACTTGTACAGTGCATGGTGACCTGCAGTGTGGTGG 690  
QY 983 TGTCCAGTTTGTCAAAATACTTGAATTCGTCACAACTTTGGCTGMAAAGAAATAGAAAG 1042  
DB 691 TGTCTCAGTTTGTCAAAATACTTGAATTCGTCACCAAGCTTTGGCTAGAAAAAG-AATAGAAG 749  
QY 1043 AAGCCACCAAGAAGCTTGGCTTCAAAATCCAGTATTTGGAGTCCATGTCAGACGACAG 1102  
DB 750 AAGCCACCAAGAAGCTTGGCTTTC-AACATCCAGTTATTGGAGTCCATGTCAGACGACAG 808  
QY 1103 ACAAGTGGGACAGAGCAGCCTTCCACCCATCGAGGAGTACATGTTACA-CGTTGAA 1161  
DB 809 ACAAGTGGGACAGAGCAGCCTTCCATCCATTTGAAGAGTACATGTTGATGTTGAA 868  
QY 1162 GAACATTTTACGTTCTCGACGAGCAATGCAAGTGGATGATAAAAAAGA 1209  
DB 869 GAACATTTTACGTTCTTTCGACGCAAAATGCAAGTGGACAAAAAGGA 916  
RESULT 10  
LOCUS BU184296  
DEFINITION AGENCOURT\_7971158 NIH\_MGC\_67 Homo sapiens cDNA clone IMAGE:6171012  
5', mRNA sequence.  
ACCESSION BU184296  
VERSION BU184296.1 GI:22698280  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 855)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM13538 Row: k Column: 13  
High quality sequence stop: 711.  
Location/Qualifiers  
1..855  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6171012"  
/tissue\_type="retinoblastoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_67"  
/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.75 kb. Library constructed by Life

BASE COUNT		265 a	172 c	224 g	193 t	1 others
ORIGIN						
Query Match						
Best Local Similarity 41.7%; Score 720.4; DB 13; Length 855;						
Matches 766; Conservative 0; Mismatches 77; Indels 0; Gaps 0;						
QY	140	TTGCAAGCTTGAACCGCTTAAACAGCAAGCAAAATGAAGACTTGAAGGCGCAATGCTGAGTCTC	199			
DB	12	TGGCAAGCTTGAACCGCTTAAACACACAGATGAAGACTTGAAGGCGAATGCCGATCTC	71			
QY	200	TCCGAATACCAAGAGCCCAATTGACACAGGGGACAGCTACAGGAAGATCCGCTGTTTAC	259			
DB	72	TCGGATACCAAGAGCCCTATTGATCAGGGCCAGCTATAGGAAGATGACGCGTTTTAG	131			
QY	260	AAGACACCTTGTAAAGCCCAAGACAGATTGAATAATACAGAAACAACTAGAAATG	319			
DB	132	AAGACACCTTGTAAAGCCCAAGACAGATTGAATAATACAGAAACAACTAGAAATG	191			
QY	320	GTCTGGGGAAGGATCATGAATCTTAAAGAGGAGATTGAAATGGAGCTAAAGAGCTCT	379			
DB	192	GTCTGGGGAAGGATCATGAATCTTAAAGAGGAGATTGAAATGGAGCTAAAGAGCTCT	251			
QY	380	GCTTTTCTACAAAGCGAACTGAAGAAATTAAGCAATTTAAGAGAAATCAACTCCAAA	439			
DB	252	GCTTTTCTACAAAGCGAACTGAAGAAATTAAGCAATTTAAGAGAAATCAACTCCAAA	311			
QY	440	GACATGACAGATGAATCTTTTGGATTAGGACACCATGAAGGCTCTATCATGACAGATC	499			
DB	312	GACATGACAGATGAATCTTTTGGATTAGGACACCATGAAGGCTCTATCATGACAGATC	371			
QY	500	TATACCTCTCAGTCAAAACAGATGAGCAGGGGATTGGCGTGAAAGAGCCAAAGATC	559			
DB	372	TATACCTCTCAGTCAAAACAGATGAGCAGGGGATTGGCGTGAAAGAGCCAAAGATC	431			
QY	560	TGACAGAGCTGGTCCAGCGGAGATAACATATCTCCAGAACTCCTAAGGACTGCAGCAAAG	619			
DB	432	TGACAGAGCTGGTCCAGCGGAGATAACATATCTCCAGAACTCCTAAGGACTGCAGCAAAG	491			
QY	620	CCAGAAAGCTGGTGTAAACATCAATAAAGGCTGGGTATGGTGTCACTCCATCAG	679			
DB	492	CCAAAGAGCTGGTGTAAATATCAAAAGGCTGGGTATGGTGTCACTCCATCATG	551			
QY	680	TGGTCTACTGTTTCAATGCTTATGCGCCAGCAACACTCATCTTGAATCTCAGA	739			
DB	552	TGGTCTACTGTTTCAATGCTTATGCGCCAGCAACACTCATCTTGAATCTCAGA	611			
QY	740	ATTGGCGCTATGCTACTGGTGGAGAGACTGTGTTTAGACCTGTAAGTGAGACATGTA	799			
DB	612	ATTGGCGCTATGCTACTGGTGGAGAGACTGTATTTAGGCGCTGTAAGTGAGACATGTA	671			
QY	800	CAGACAGATCTGGCTCTCCACTGGACACTGGTCAAGGTGAAGTAAATGACAAAACATTC	859			
DB	672	AAGACAGATCTGGCATCTCCACTGGACACTGGTCAAGGTGAAGTGAAGGACAAAATGTC	731			
QY	860	AAGTGGTCAAGTCCCACTTGTAGACAGCTCCATCTCCGCGCTCTTACTTACCACTGG	919			
DB	732	AAGTGGTCAAGTCCCACTTGTAGACAGCTCTTCACTCCGCGCTCTTACTTACCACTGG	791			
QY	920	CTGTTCCAGAGAGCTTCAGACAGCTCCCTAAGAGTCCATGGTGACCTGCAGTGGT	979			
DB	792	CTGTACCAAGAGACCTCCAGATCGACTTGTACAGTGCATGGTGACCTGGCCNGGTGG	851			
QY	980	GGG 982				
DB	852	GGG 854				
RESULT 11						
BU513162						
LOCUS						
DEFINITION						
AGENCOURT_10116686 NIH_MGC_134 Mus musculus cDNA clone						
EST 12-SEP-2002						
linear						
mRNA						
937 bp						

IMAGE:6508154 5', mRNA sequence.	
ACCESSION	BU513162
VERSION	BU513162.1
KEYWORDS	EST.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 937)
TITLE	NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished Contact: Robert Strausberg, Ph.D. Email: c9abps-remail.nih.gov Tissue Procurement: Dr. David Rowe cDNA Library Preparation: Invitrogen Corp DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM14073 row: g column: 03 High quality sequence stop: 751. Location/Qualifiers 1. .937 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="IMAGE:6508154" /tissue_type="undifferentiated limb" /lab_host="DH10B (phage-resistant)" /clone_lib="NIH_MGC_134" /note="Vector: pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1-7 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library." BASE COUNT 260 a 220 c 204 g 250 t 3 others ORIGIN Query Match 40.0%; Score 704.6; DB 13; Length 937; Best Local Similarity 91.4%; Pred. No. 7.8e-172; Matches 819; Conservative 0; Mismatches 5; Indels 72; Gaps 4 QY 690 TTTTCATGATTGCTTATGGCAACCCAGCGAACACTCATCTTGGAAATCTCAGAAATGGCGCTA 749 DB 1 TTTTCATGATTGCTTATGGCAACCCAGCGAACACTCATCTTGGAAATCTCAGAAATGGCGCTA 60 QY 750 TGTCTACTGTGATGATGAGACTGTGTTTAGACTCTGTAGTGTAGACATGTAGACAGATC 809 DB 61 TGTCTACTGTGATGATGAGACTGTGTTTAGACTCTGTAGACTCTTAAGTGTAGACATGTAGACAGATC 120 QY 810 TGGCTCTCCACTGGNACTGTGAGTGAAGTAAATGACAAAACATTTCAAGTGTCTGA 869 DB 121 TGGCTCTCCACTGGNACTGTGAGTGAAGTAAATGACAAAACATTTCAAGTGTCTGA 180 QY 870 GTCTCCCACTGTAGATGAGCTCCATCTCGGCTCTCTTACTTACCACCTGGCTGTCTCCAGA 929 DB 181 GCTCCCACTGTAGATGAGCTCCATCTCGGCTCTCTTACTTACCACCTGGCTGTCTCCAGA 240 QY 930 AGACCTTGCAGACCGACTCCTTAAGAGTCCATGGTGAACCTTGCAGTGTGGTGTGCCA 989 DB 241 AGACCTTGCAGACCGACTCCTTAAGAGTCCATGGTGAACCTTGCAGTGTGGTGTGCCA 300 QY 990 GTTGTCAAACTACTTATTCGTCACCACTTGGCTGGAAAAGGAATAAGAAAGCCAC 104 DB 301 GTTGTCAAACTACTTATTCGTCACCACTTGGCTGGAAAAGGAATAAGAAAGCCAC 360 QY 1050 CAAGAAGCTTGGCTTCAAACTCCAGTTATTCGAGTCCATGTTCAGACGACACAGACAAAGT 1109 DB 361 CAAGAAGCTTGGCTTCAAACTCCAGTTATTCGAGTCCATGTTCAGACGACACAGACAAAGT 420 QY 1110 GGGAAACAGAAAGCAGCTTCCACCCCATCGAGGATCATGGTACATGGTGAAGAACATTT 1169



Db 421 GGAACAGAGAGAGCGCTTCCACCCCATCGAGGAGTACATGGTACAGTGTGAAGAACATTT 480  
QY 1170 TCAGCTTCTTCGACGACGAGATGCAAGTGGATAAAAAGAGATATATCTGGCTACTGATGA 1229  
Db 481 TCAGCTTCTTCGACGACGAGAT----- 501  
QY 1230 TCCTACTTTGTTAAAGGAGGCAAGACAAAGTACTCCAATTATGAAATTTATTTAGTGATAA 1289  
Db 502 -----GTACTCCAATTATGAAATTTATTTAGTGATAA 531  
QY 1290 CTCTATTTCTTGGTACGCTGGACTACAAATCGGTACACAGAAATTCACCTCGGGGTGT 1349  
Db 532 CTCTATTTCTTGGTACGCTGGACTACAAATCGGTACACAGAAATTCACCTCGGGGTGT 591  
QY 1350 GATCCTGGATATACACTTCTCTCAGAGCTGACTTTCTAGTGTGTACTTTTTCATCCCA 1409  
Db 592 GATCCTGGATATACACTTCTCTCAGAGCTGACTTTCTAGTGTGTACTTTTTCATCCCA 651  
QY 1410 GGTCTGTGGTGTCTTATGAAATCATGCAAAACCTGCAATCTGATGCTCTGCGAACTT 1469  
Db 652 GGTCTGTGGTGTCTTATGAAATCATGCAAAACCTGCAATCTGATGCTCTGCGAACTT 711  
QY 1470 CAATCTTTGGATGACATCTACTATTTGGAGGCCAAATGCCCAATCAGATTTGCTGT 1529  
Db 712 -CAATCTTTGGATGACATCTACTATTTGGAGGCCAAATGCCCAATCAGATTTGCTGT 770  
QY 1530 TTATCTCTCAAAACC-TCGAACTGAAGA-GGAAATTCGAATGGAACCTGGAGATAT 1583  
Db 771 TTATCTCTCAAAACCCTTGAAGAGGAGGAAATTCGAATGGAACCTGGAGAT 826

## RESULT 12

BX450010

LOCUS

DEFINITION BX450010 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA clone CS0DG005Y007 5-PRIME, mRNA sequence.

ACCESSION BX450010

VERSION BX450010.1

KEYWORDS EST.

SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1. (bases 1 to 1027)

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished

Contact: Genoscope

Genoscope - Centre National de Sequencage

Bp 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 7499.f For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0AS0092B08P1&amp;cluster=7499.f. Contact :

Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Paradise Avenue Genoscope sequence ID : CS0AS0092B08QP1.

Location/Qualifiers

1. .1027

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0DG005Y007"

/tissue\_type="B CELLS (RAMOS CELL LINE)"

/cell\_line="RAMOS CELL LINE"

/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed

with a NotI-oligo(dT) primer. Five prime end enriched,

double-strand cDNA was digested with Not I and cloned into

the Not I and EcoRV sites of the pCMVSPORT 6 vector.

Library was not normalized."

## FEATURES

source

BASE COUNT 320 a 188 c 258 g 252 t 9 others  
ORIGIN  
Query Match 40.3%; Score 696.8; DB 13; Length 1027;  
Best Local Similarity 92.4%; Pred. No. 8.6e-170;  
Matches 728; Conservative 4; Mismatches 56; Indels 0; Gaps 0;  
QY 1 ATGCGGCATGAGACTGGTTCCTGGCGTTGGATTATGCTCATTTCTTTTCCCTGGGGGACC 60  
Db 240 ATGCGGCATGAGACTGGTTCCTGGCGTTGGATTATGCTCATTTCTTTTCCCTGGGGGACC 299  
QY 61 TTGTTATTTATATAGTGGTCAATTTGCTTCGAGATATGACCACTCTGATCACTCCAGC 120  
Db 300 TTGCTGTTTTATATAGTGGTCAATTTGCTTCGAGATATGACCACTCTGATCACTCTAGC 359  
QY 121 AGAGAACTCTCCAAAGATTCTTGCAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 180  
Db 360 CGAGAACTGTTCCAAGATTCTGCAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 419  
QY 181 AGGCGAATGGTGTAGTCTCTCCGAATACCAAGAGGCCCATTTGACAGGGGACAGCTACA 240  
Db 420 AGGCGAATGGGCGAATCTCTCCGAATACCAAGAGGCCCATTTGATGATGAGGGGCCAGCTATA 479  
QY 241 GGAAGAGTCCGTGTTTTAGAGAGACAGCTTGAAGCCAAAGACAGATTGAAAATTTAC 300  
Db 480 GGAAGAGTCCGTGTTTTAGAGAGACAGCTTGAAGCCAAAGACAGATTGAAAATTTAC 539  
QY 301 AAGAAAACAAGCTAGAAATGTTCTGGGGAAGGATCATGAAATCTTTAAGAGGAGGATTGAA 360  
Db 540 AAGAAAACAAGCTAGAAATGTTCTGGGGAAGGATCATGAAATCTTTAAGAGGAGGATTGAA 599  
QY 361 AATGAGCTAAGAGCTCTGTTTTTCTCAAGCGCAACTGGAAGAAATTAAGCAATTTA 420  
Db 600 AATGAGCTAAGAGCTCTGTTTTTCTCAAGCGCAACTGGAAGAAATTAAGCAATTTA 659  
QY 421 GAAAGAAATGAATCCCAAGACATGCAGATGAAATCTTTTGGATTTTGAAGACACCATGAA 480  
Db 660 GAAAGAAATGAATCCCAAGACATGCAGATGAAATCTTTTGGATTTTGAAGACATCATGAA 719  
QY 481 AGGTCTATCATGACAGATCTATCTACTCTCAAGCAAGCATGGAGGAGGATTGGCGT 540  
Db 720 AGGTCTATATGACGATCTATCTACTCTCAAGCAAGCATGGAGGAGGATTGGCGG 779  
QY 541 GAAAGAGGCGCAAGATCTCAGAGCTGTGTCCAGCGGAGAAATTAATATCTCCAGAAAT 600  
Db 780 GAAAGAGGCGCAAGATCTCAGAGCTGTGTCCAGCGGAGAAATTAATATCTCCAGAAAT 839  
QY 601 CCTAAGGACTGCAGCAAGCCAGGAAGCTGTGTGTAACATCAATATAAGGCTGTGGCTAT 660  
Db 840 CCTAAGGACTGCAGCAAGCCAGGAAGCTGTGTGTAACATCAATATAAGGCTGTGGCTAT 899  
QY 661 GGTGTCAACTCCATCAGCTGTCTACTGTGTTTCATGATTTGTCATGTCATATGGCACCAGCAACA 720  
Db 900 GGTGTCAACTCCATCAGCTGTCTACTGTGTTTCATGATTTGTCATATGGCACCAGCAACA 959  
QY 721 CTCAATCTTGAATCTCAGAAATTTGGCGCTATCTCTACTGTGTGATGGAGACTGTGTTAGA 780  
Db 960 CTCAATCTTGAATCTCAGAAATTTGGCGCTATCTCTACTGTGTGATGGAGACTGTGTTAGA 1019  
QY 781 CCTGTAG 788  
Db 1020 TCTGTAG 1027

## RESULT 13

AUI24128

LOCUS

DEFINITION AUI24128 NT2RM2 Homo sapiens cDNA clone NT2RM2001714 5', mRNA

ACCESSION AUI24128

VERSION AUI24128.1

KEYWORDS EST.

SOURCE Homo sapiens (human)



ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 834)  
AUTHORS Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S., Saito, K., Yamamoto, J.,  
Nakamura, Y., Nishikawa, T., Nagai, T., Suzuki, Y., Sugano, S. and  
Isogai, T.  
TITLE HRI human cDNA project (Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S.,  
Saito, K., Yamamoto, J., Nakamura, Y., Nishikawa, T., Nagai, T., Suzuki,  
Y., Sugano, S., Isogai, T.)  
JOURNAL Unpublished  
COMMENT Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975  
Fax: 81-438-52-3986  
Email: genomics@hri.co.jp  
HRI human cDNA project; 5' - & 3' - end one pass sequencing; Helix  
Research Institute; cDNA library construction; Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.  
FEATURES  
Location/Qualifiers  
1..834  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NT2RM2001714"  
/cell\_type="teratocarcinoma"  
/cell\_lines="NT2"  
/clone\_lib="NT2RM2"  
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal  
precursor cells"  
BASE COUNT 249 a 187 c 181 g 215 t 2 others  
ORIGIN  
Query Match 39.8%; Score 687.4; DB 9; Length 834;  
Best Local Similarity 90.4%; Pred. No. 2.2e-167;  
Matches 744; Conservative 0; Mismatches 78; Indels 1; Gaps 1;  
QY 905 CTACTTACCTGGCTGTTCCAGAAAGACCTTGCAGACCGACTCTCAGAGTCCATGATGTTG 964  
DB 1 CATATTTACCTTGGCTGTACAGAAAGACCTCGCAGATCGACTTGTACGAGTGCATGTTG 60  
QY 965 ACCCTGCAGTGTGGGTGTCCTCCAGTTGTCATATCTGATTCCTCCACAACTTCGCG 1024  
DB 61 ACCCTGCAGTGTGGGTGTCCTCCAGTTGTCATATCTGATTCCTCCACAACTTCGCG 120  
QY 1025 TGGAAAGAAATAGAAAGACCCACCAAGAGCTTGGCTTCAAAATCCAGTTATTGGAG 1084  
DB 121 TAGAAAGAAATAGAAAGACCCACCAAGAGCTTGGCTTCAAAATCCAGTTATTGGAG 180  
QY 1085 TCCATGTCCAGCGCAGACAAAGTGGGAACAGAGCGCTTCCACCCCATCGAGAGT 1144  
DB 181 TCCATGTCCAGCGCAGACAAAGTGGGAACAGAGCGCTTCCATCCCATTTGAAGAGT 240  
QY 1145 ACATGGTACAGTTGAAGAACATTTTCAGTCTCCGACGAGATGCAAGTGGATATAA 1204  
DB 241 ACATGGTACAGTTGAAGAACATTTTCAGTCTCCGACGAGATGCAAGTGGCAAAA 300  
QY 1205 AAAGAGTATATCGCTACTGATGATCTTCTTTAAAGAGGCAAGACAAAGTACT 1264  
DB 301 AAAGAGTATATCGCTACTGATGATCTTCTTTAAAGAGGCAAGACAAAGTACT 360  
QY 1265 CCAATTATGAATTTATTAGTGATACTCTATTCTTGGTCAGCTGGACTACCAATCGGT 1324  
DB 361 CCAATTATGAATTTATTAGTGATACTCTATTCTTGGTCAGCTGGACTGGCAATCGAT 420  
QY 1325 ACACAGAAAATTCACCTTCGGGTGTGATCTCGGATATACATTTCTCTCACAGGCTGACT 1384  
DB 421 ACACAGAAAATTCACCTTCGGGTGTGATCTCGGATATACATTTCTCTCACAGGCTGACT 480  
QY 1385 TTCTAGTGTGACTTTTTCATCCCGAGTCTGTCGGGTGCTTATGAATCATGCAAAACC 1444

Db 481 TCCTAGTGTGACTTTTTCATCCAGGCTCTGTCAGTGTCTATGAAATATGCAAAACAC 540  
QY 1445 TGCATCTGATGCTCTGCGAACTTCCATTCTTTGGATGACATCTACTATTTCGGAGGCC 1504  
Db 541 TAGTCTCTGATGCTCTGCGAACTTCCATTCTTTAGATGACATCTACTATTTCGGGGCC 600  
QY 1505 AAAATGCCCAATCACTAGATGCTGTTTATCCTCAAAACCTCGAACTGGAAGAGAAATTC 1564  
Db 601 AGAATGCCCAATCAATTAATTTGCCATTTATGCTCACCAACCCCGAATCGCATGAAATTC 660  
QY 1565 CAATGGAACCTGGAGATATCATTTGGTGTGCTGGAACCATTTGGATGGTTATTCTAAAG 1624  
Db 661 CCATGGAACCTGGAGATATCATTTGGTGTGCTGGAATCATTTGGATGGCTATTCCAAAG 720  
QY 1625 GTATCAACAGAAAACCTTGGAAAACAGGCTTATATCCTCTCACAAGTCCGAGAGAAGA 1684  
Db 721 GTGTCAACAGAAAATTTGGGAAGACGCGCTATAT-CTCTACAAAGTTCCGAGAGAAGA 779  
QY 1685 TAGAAACAGTCAAGTATCCACATATCTGAAGCTGAAAAATA 1727  
Db 780 TAGAAACGTCAGTACCCACATATCTCGAGCTGAGAAATA 822  
RESULT 14  
BG973221 702 bp mRNA linear EST 12-JUN-2001  
LOCUS 602842563f1 NCI\_CGAP\_Mam4 Mus musculus cDNA clone IMAGE:4978383 5',  
DEFINITION mRNA sequence.  
ACCESSION BG973221  
VERSION BG973221.1 GI:14360858  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 702)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth  
Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10974 row: n column: 16  
High quality sequence stop: 694.  
FEATURES  
Location/Qualifiers  
1..702  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="NMRI"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4978383"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/clone\_lib="NCI\_CGAP Mam4"  
/note="Oxgan: mammary; Vector: pCMV-SPORT6; Site 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigators  
providing samples: Lothar Hennighausen/Priscilla Furth,  
NIH Reference for transgenic model: Li et al., Cell Growth  
and Differentiation 7, 3-11 (1996)."  
BASE COUNT 217 a 145 c 174 g 166 t  
ORIGIN  
Query Match 38.3%; Score 660.2; DB 12; Length 702;

Qy	846	TGACAAAAACATTTCAGTGGTGCAGCTCCCACTTGTAGACAGCTCCATCTCTGGGCTCC	905
Db	61	TGCAAAAAACATTTCAGTGGTGCAGCTCCCACTTGTAGACAGCTCCATCTCTGGGCTCC	120
Qy	906	TTATCTTACACATCGGTGTTTCAGAGACCTTCGCAGCCGACTCTTAAGAGTCCATGGTGA	965
Db	121	TTACTTTACCACTGGCTGTTTCCAGAGACCTTCGCAGCCGACTCTTAAGAGTCCATGGTGA	180
Qy	966	CCCTCGAGTGGTGGGTGCTCCAGTTTCGTCAAAATCTGATTTCGTCACAACTTGGCT	1025
Db	181	CCCTCGAGTGGTGGGTGCTCCAGTTTCGTCAAAATCTGATTTCGTCACAACTTGGCT	240

241	GGAAGGAAATAGAAAGCCACCAAGAGCTTGGCTTCAACATCCAGTTATTGGAGT	300
Db		
1086	CCATGT'CAGACGCACAGACAAAGTGGGAACAGAAAGCAGCGCTTCCACCCCATCGAGAGTA	1145
Qy		
301	CCATGT'CAGACGCACAGACAAAGTGGGAACAGAAAGCAGCGCTTCCACCCCATCGAGAGTA	360
Db		
1146	CATGTTACACGTTGAAGACATTTTCAGCTTCTCGCACGCAGAAATGCAAGTGGATAAAAA	1205
Qy		
361	CATGTTACACGTTGAAGACATTTTCAGCTTCTCGCACGCAGAAATGCAAGTGGATAAAAA	420
Db		
1206	AAGAGTATATCTGGCTACTGATGATCCTACTTTGTTTAAAGGAGGCAAGCAAGTACTC	1265
Qy		
421	AAGAGTATATCTGGCTACTGATGATCCTACTTTGTTTAAAGGAGGCAAGCAAGTACTC	480
Db		
1266	CAATTATGAATTTATTAGTGTAACTCTATTTCCTGGTCAGCTGGAGTACACAAATCGGTA	1325
Qy		
481	CAATTATGAATTTATTAGTGTAACTCTATTTCCTGGTCAGCTGGAGTACACAAATCGGTA	540
Db		
1326	CAAGAAAAATTCATCTGGGGGTGTGATCTCGGATATACACTTTTCTCTCAGGCTGACTT	1385
Qy		
541	CACAGAAAAATTCATCTGGGGGTGTGATCTCGGATATACACTTTTCTCTCAGGCTGACTT	600
Db		



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 08:39:39 ; Search time 38.5 Seconds  
(without alignments)

2370.592 Million cell updates/sec

Title: US-09-971-773-23

Perfect score: 3081

Sequence: 1 MRAWTGSWRWIMLILFAWGT.....YKREKTIETVKYPTYPEAEK 575

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3081	100.0	575	23 ABG34135	Antibody productio
2	3024	98.1	575	23 ABG34136	Antibody productio
3	2984	96.9	575	18 AAW22125	Human alpha 1-6 fu
4	2884	96.9	575	23 ABB08405	Alpha1,6-fucosyl t
5	2885	93.6	575	18 AAW22124	Pig alpha 1-6 fuco
6	2639	85.7	515	22 AAG73884	Human colon cancer
7	1771	57.5	339	22 AAB75061	Human alpha 1-6 fu
8	1489.5	48.3	619	22 ABB59117	Drosophila melanog
9	1193	38.7	233	22 AAB75062	Human alpha 1-6 fu

10	431	14.0	82	22	ABG48804	Human liver peptid
11	431	14.0	82	22	ABB28797	Peptide #1448 enco
12	431	14.0	82	22	ABB33985	Peptide #1491 enco
13	431	14.0	82	22	ABB19422	Protein #1421 enco
14	431	14.0	82	22	AAW54747	Human brain expres
15	431	14.0	82	22	AAW67139	Human bone marrow
16	431	14.0	82	22	AAW15003	Peptide #1437 enco
17	431	14.0	82	22	AAW27440	Peptide #1477 enco
18	431	14.0	82	22	AAW02730	Peptide #1412 enco
19	431	14.0	82	23	ABG36801	Human peptide enco
20	425	13.8	81	22	AAW64859	Human brain expres
21	425	13.8	81	22	AAW37789	Peptide #11826 enc
22	133	4.3	1959	23	ABJ10604	Human novel protei
23	131	4.3	25	18	AAW22131	Fragment #2 of hum
24	130.5	4.2	751	22	ABW62477	Drosophila melanog
25	124.5	4.0	1960	22	AAW78854	Human protein SEQ
26	124.5	4.0	2143	22	ABG01716	Novel human diagno
27	124	4.0	1025	23	ABB90967	Herbicide activ
28	124	4.0	1285	24	ABR47540	Breast cancer asso
29	124	4.0	2442	21	AAW77575	Human cytoskeletal
30	123	4.0	1963	22	AAW79838	Human protein SEQ
31	121.5	3.9	898	24	AAE33671	Human structural a
32	121.5	3.9	1014	23	ABG61923	Prostate cancer-as
33	120.5	3.9	3542	22	AAW62142	P. faicparum FCR3
34	119.5	3.9	612	22	AAW95546	Human protein sequ
35	118.5	3.8	976	22	AAW66581	Human SCP-1 mutain
36	118.5	3.8	976	24	ABP74709	Human SCP-1 protei
37	117.5	3.8	1427	12	AAW10534	Human 160kD mediat
38	116.5	3.8	931	22	AAW79504	Human protein SEQ
39	116.5	3.8	990	22	AAW78520	Human protein SEQ
40	116	3.8	214	18	AAW09037	Osteoclast stimula
41	116	3.8	214	18	AAW12706	Osteoclast stimula
42	116	3.8	229	17	AAW05403	Human clone 5 prot
43	116	3.8	283	21	AAW28744	Arabidopsis thalia
44	116	3.8	2383	23	ABG56531	Human breast speci
45	115.5	3.7	888	23	AAW83013	Human homologue of

#### ALIGNMENTS

RESULT 1

ABG34135

ID ABG34135 standard; Protein; 575 AA.

AC ABG34135;

XX

DT 15-JUL-2002 (first entry)

XX

DE Antibody production method related protein #1.

XX

KW Antibody production; cytostatic; immunomodulator; vasotropic; virucide;  
KW antibacterial; antinflammatory; antiallergic; allergy; inflammation;  
KW autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO;  
KW tumour; circulatory disease; infection.

XX

OS Cricetulus griseus.

XX

PN WO200231140-A1.

XX

PD 18-APR-2002.

XX

PF 05-OCT-2001; 2001WO-JP08804.

XX

PR 06-OCT-2000; 2000JP-0308526.

XX

PA (KYOWA) KYOWA HAKKO KOGYO KK.

XX

PI Kanda Y, Sato M, Nakamura K, Uchida K, Shinkawa T, Yamane N;

PI Hosaka E, Yamano K, Yamasaki M, Hanai N;

XX

DR WPI; 2002-340182/37.

XX

PT Cells producing antibody compositions including antibody fragments and  
PT fusion proteins with Fc domain of antibody, useful for prevention or  
PT treatment of cancer, immune diseases, circulatory diseases and  
PT infections -  
XX  
XX  
PS Claim 23; Page 18-21; 314pp; Japanese.  
XX  
CC This invention relates to novel method for antibody production  
CC comprising a Chinese hamster ovarian tissue-originated (CHO) cell  
CC transferred with a gene encoding an antibody molecule for producing a  
CC composition comprising an antibody molecule with an Fc domain bonded  
CC to the N-glycoside linkage complex sugar chain. The produced antibody  
CC compositions are drugs for prevention or treatment of diseases  
CC accompanying tumour, allergy or inflammation, autoimmune diseases,  
CC circulatory diseases, and viral and bacterial infections. The  
CC antibodies can be stably produced using the method of the invention  
CC with high binding activity and potency thus leading to high safety and  
CC reduced side effects when applied alone or in combination with other  
CC drugs for therapy. The present sequence represents a protein  
CC molecule used in the method of the invention.  
XX  
XX  
SQ Sequence 575 AA;  
Query Match 100.0%; Score 3081; DB 23; Length 575;  
Best Local Similarity 100.0%; Pred. No. 5.5e-272;  
Matches 575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MRAWTGSRWIMLILFAWGTLIFYGGHLVRNDHPDHSRSLSKILAKLERLKQONEDL 60  
DB 1 MRAWTGSRWIMLILFAWGTLIFYGGHLVRNDHPDHSRSLSKILAKLERLKQONEDL 60  
QY 61 RMAESLRIPGPIDOGTATGRVRLVEQLVKAKEQIENYKQARNDLGKDHILRRRIE 120  
DB 61 RMAESLRIPGPIDOGTATGRVRLVEQLVKAKEQIENYKQARNDLGKDHILRRRIE 120  
QY 121 NGAKELWFFLQSELKKLKEGNEQLQRADEILDLGHHERSIMTDLVYLSQTDGAGWR 180  
DB 121 NGAKELWFFLQSELKKLKEGNEQLQRADEILDLGHHERSIMTDLVYLSQTDGAGWR 180  
QY 181 EKEADLTTELVRRTYILQNPDKCSKARKLVNKGCGYCOLHHVYCFMAYGTORT 240  
DB 181 EKEADLTTELVRRTYILQNPDKCSKARKLVNKGCGYCOLHHVYCFMAYGTORT 240  
QY 241 LILESQNRWYATGGWETVFRPVSETCTDRSGLSTGHWSGEVKDKNVQVVELPIVDSLHPR 300  
DB 241 LILESQNRWYATGGWETVFRPVSETCTDRSGLSTGHWSGEVKDKNVQVVELPIVDSLHPR 300  
QY 301 PPYPLAVPEDLADRLRLRVHGDPAVWVVSQFVKYLIRPQWLERIEETTKLGFKEHVI 360  
DB 301 PPYPLAVPEDLADRLRLRVHGDPAVWVVSQFVKYLIRPQWLERIEETTKLGFKEHVI 360  
QY 361 GVHVRTDKVGEAAFPHEEYVWVEHFQLLERMMKVDKRVYLATDDPSLLKEATK 420  
DB 361 GVHVRTDKVGEAAFPHEEYVWVEHFQLLERMMKVDKRVYLATDDPSLLKEATK 420  
QY 421 YSNYEFISDNSISWSAGLHNRYTENSRLGVILDIHFLSQADFLVCTFSSQCRVAYEIMQ 480  
DB 421 YSNYEFISDNSISWSAGLHNRYTENSRLGVILDIHFLSQADFLVCTFSSQCRVAYEIMQ 480  
QY 481 TLHPDASANFHLDDIYFPGQNAHQIAVYPHQPTKEEIPMEPGDIIIGVAGNHNHNGYS 540  
DB 481 TLHPDASANFHLDDIYFPGQNAHQIAVYPHQPTKEEIPMEPGDIIIGVAGNHNHNGYS 540  
QY 541 KGVNRKLGKTLGVPSYKREKLETVKYPTYPEAK 575  
DB 541 KGVNRKLGKTLGVPSYKREKLETVKYPTYPEAK 575

RESULT 2  
ABG34136  
ID ABG34136 standard; Protein; 575 AA.  
XX  
AC ABG34136;

XX 15-JUL-2002 (first entry)  
XX  
DE Antibody production method related protein #1.  
XX  
KW Antibody production; cytostatic; immunomodulator; vasotropic; virucide;  
KW antibacterial; antiinflammatory; antiallergic; allergy; inflammation;  
KW autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO;  
XX tumour; circulatory disease; infection.  
OS Mus musculus.  
XX WO200231140-A1.  
XX 18-APR-2002.  
XX 05-OCT-2001; 2001WO-JP08804.  
XX 06-OCT-2000; 2000JP-0308526.  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX Kanda Y, Satoh M, Nakamura K, Uchida K, Shinkawa T, Yamane N;  
PI Hosaka E, Yamano K, Yamasaki M, Hanai N;  
XX WPI; 2002-340182/37.  
XX  
PT Cells producing antibody compositions including antibody fragments and  
PT fusion proteins with Fc domain of antibody, useful for prevention or  
PT treatment of cancer, immune diseases, circulatory diseases and  
PT infections -  
XX  
XX  
PS Claim 33; Page 21-23; 314pp; Japanese.  
CC  
CC This invention relates to novel method for antibody production  
CC comprising a Chinese hamster ovarian tissue-originated (CHO) cell  
CC transferred with a gene encoding an antibody molecule for producing a  
CC composition comprising an antibody molecule with an Fc domain bonded  
CC to the N-glycoside linkage complex sugar chain. The produced antibody  
CC compositions are drugs for prevention or treatment of diseases  
CC accompanying tumour, allergy or inflammation, autoimmune diseases,  
CC circulatory diseases, and viral and bacterial infections. The  
CC antibodies can be stably produced using the method of the invention  
CC with high binding activity and potency thus leading to high safety and  
CC reduced side effects when applied alone or in combination with other  
CC drugs for therapy. The present sequence represents a protein  
CC molecule used in the method of the invention.  
XX  
XX  
SQ Sequence 575 AA;  
Query Match 98.1%; Score 3024; DB 23; Length 575;  
Best Local Similarity 97.6%; Pred. No. 8.8e-267;  
Matches 561; Conservative 9; Mismatches 5; Indels 0; Gaps 0;  
QY 1 MRAWTGSRWIMLILFAWGTLIFYGGHLVRNDHPDHSRSLSKILAKLERLKQONEDL 60  
DB 1 MRAWTGSRWIMLILFAWGTLIFYGGHLVRNDHPDHSRSLSKILAKLERLKQONEDL 60  
QY 61 RMAESLRIPGPIDOGTATGRVRLVEQLVKAKEQIENYKQARNDLGKDHILRRRIE 120  
DB 61 RMAESLRIPGPIDOGTATGRVRLVEQLVKAKEQIENYKQARNDLGKDHILRRRIE 120  
QY 121 NGAKELWFFLQSELKKLKEGNEQLQRADEILDLGHHERSIMTDLVYLSQTDGAGWR 180  
DB 121 NGAKELWFFLQSELKKLKEGNEQLQRADEILDLGHHERSIMTDLVYLSQTDGAGWR 180  
QY 181 EKEADLTTELVRRTYILQNPDKCSKARKLVNKGCGYCOLHHVYCFMAYGTORT 240  
DB 181 EKEADLTTELVRRTYILQNPDKCSKARKLVNKGCGYCOLHHVYCFMAYGTORT 240  
QY 241 LILESQNRWYATGGWETVFRPVSETCTDRSGLSTGHWSGEVKDKNVQVVELPIVDSLHPR 300  
DB 241 LILESQNRWYATGGWETVFRPVSETCTDRSGLSTGHWSGEVKDKNVQVVELPIVDSLHPR 300

QY 301 PPYLPLAVPEDLADLLRVHGDPAVWVWSQFVKYLIRPQWLERIEETTKLGFKEHPVI 360  
 DB 301 PPYLPLAVPEDLADLLRVHGDPAVWVWSQFVKYLIRPQWLERIEETTKLGFKEHPVI 360  
 QY 361 GVHVRTDKVGTAAAFHPFIEEYMHVBEHFQLLERRMKVDKRVYLATDTPSLKEAKTK 420  
 DB 361 GVHVRTDKVGTAAAFHPFIEEYMHVBEHFQLLERRMKVDKRVYLATDTPSLKEAKTK 420  
 QY 421 YSNYEFISDINSISWAGLHNRNTENSLRGVILDIHFLSQADPLVCTFSSQVCRVAYEIMQ 480  
 DB 421 YSNYEFISDINSISWAGLHNRNTENSLRGVILDIHFLSQADPLVCTFSSQVCRVAYEIMQ 480  
 QY 481 TLHPDASANFSLDDIYFVGQNAHNOIAVYPHOPRTKEEIPMEPGDIIGVAGNHNGYS 540  
 DB 481 TLHPDASANFSLDDIYFVGQNAHNOIAVYPHOPRTKEEIPMEPGDIIGVAGNHNGYS 540  
 QY 541 KGVNRKLGKTGLYPSYKVKREKIEIVKYPTYPEAEK 575  
 DB 541 KGVNRKLGKTGLYPSYKVKREKIEIVKYPTYPEAEK 575

## RESULT 3

AAW22125  
 ID AAW22125 standard; Protein; 575 AA.

AC AAW22125;

DT 05-MAR-1998 (first entry)

DE Human alpha 1-6 fucosyltransferase.

KW Alpha 1-6 fucosyltransferase; enzyme; pig; human; fucose transfer;  
 KW Guanosine diphosphate; sugar chain synthesis; modification; antibody;  
 KW GINAC; cancer diagnosis.

OS Homo sapiens.

PN WO9727303-A1.

PD 31-JUL-1997.

PF 23-JAN-1997; 97WO-JP00171.

PR 22-JUL-1996; 96JP-0192260.

PR 24-JAN-1996; 96JP-0010365.

PR 21-JUN-1996; 96JP-0161648.

PR 24-JUN-1996; 96JP-0162813.

PA (TOYM) TOYO BOSEKI KK.

PI Shiba T, Taniguchi N, Uozumi N, Yanagidani S;

DR WPI; 1997-393690/36.

DR N-PSDB; AAT76574.

XX Human or pig alpha 1-6 fucosyltransferase and DNA encoding it - for  
 PT synthesis and modification of sugar chains and used as an antigen  
 PT for production of diagnostic antibodies

XX Claim 17; Page 39-43; 61pp; Japanese.

XX AAW22124 and AAW22125 represent the pig and human alpha 1-6  
 CC fucosyltransferases of the invention, respectively. The enzyme transfers  
 CC fucose from guanosine diphosphate to the 6-hydroxyl group of the GlcNAc  
 CC nearest to R in the receptor molecule: (GlcNAc-beta 1-2-Man-alpha 1-6)  
 CC (GlcNAc-beta 1-2-Man-alpha 1-3)-Man-beta 1-4-GlcNAc-R to give  
 CC (GlcNAc-beta 1-2-Man-alpha 1-6)(GlcNAc-beta 1-2-Man-alpha 1-3)-Man-beta  
 CC 1-4-GlcNAc-beta 1-4(Fuc-alpha 1-6)(GlcNAc-R). It has an optimum pH of about  
 CC 7.0 (pig) or 7.5 (human), and is stable over the pH range 4-10 after 5  
 CC hours at 4 degrees C. The optimum working temperature of the  
 CC alpha 1-6 fucosyltransferases is 30-37 degrees C. A bivalent metal is  
 CC not required for activity of the enzyme, and the enzyme is not inhibited

CC in the presence of 5 mM EDTA. The enzyme is useful in the synthesis and  
 CC modification of sugar chains, and as antigen for the production of  
 CC antibodies recognising the enzyme. The antibodies can be used for the  
 CC diagnosis of cancer and other diseases.

XX SQ Sequence 575 AA;

Query Match 96.9%; Score 2984; DB 18; Length 575;

Best Local Similarity 96.2%; Pred. No. 3.9e-263;

Matches 553; Conservative 10; Mismatches 12; Indels 0; Gaps 0;

QY 1 MRVATGSRWIMLILFWAGTLLFYIGHLVRDNDHPDHSSELKILAKLERLQQNEDL 60

DB 1 MRPTGSRWIMLILFWAGTLLFYIGHLVRDNDHPDHSSELKILAKLERLQQNEDL 60

QY 61 RMAESLRIPGPDIDGTATGRVRLVEQLVKAQEIQENYKQARNDLGKDEILRRRIE 120

DB 61 RMAESLRIPGPDIDGTATGRVRLVEQLVKAQEIQENYKQARNDLGKDEILRRRIE 120

QY 121 NGAKELWFFLQSELKILKLEGNELQRLHADRILLDLGHHERSIMTDLVYLSQTDGAGWR 180

DB 121 NGAKELWFFLQSELKILKLEGNELQRLHADRILLDLGHHERSIMTDLVYLSQTDGAGWR 180

QY 181 EKEAKDLTELVRRIITYLQNPDCSKARKLVNINKGCGYCOLHHVVCFMIAYGTOPT 240

DB 181 EKEAKDLTELVRRIITYLQNPDCSKARKLVNINKGCGYCOLHHVVCFMIAYGTOPT 240

QY 241 LILESQNRVATGNETVFRPVSETCTDRSLSTGHWSGEVKDNQVQVELPIVDSLHPR 300

DB 241 LILESQNRVATGNETVFRPVSETCTDRSLSTGHWSGEVKDNQVQVELPIVDSLHPR 300

QY 301 PPYLPLAVPEDLADLLRVHGDPAVWVWSQFVKYLIRPQWLERIEETTKLGFKEHPVI 360

DB 301 PPYLPLAVPEDLADLLRVHGDPAVWVWSQFVKYLIRPQWLERIEETTKLGFKEHPVI 360

QY 361 GVHVRTDKVGTAAAFHPFIEEYMHVBEHFQLLERRMKVDKRVYLATDTPSLKEAKTK 420

DB 361 GVHVRTDKVGTAAAFHPFIEEYMHVBEHFQLLERRMKVDKRVYLATDTPSLKEAKTK 420

QY 421 YSNYEFISDINSISWAGLHNRNTENSLRGVILDIHFLSQADPLVCTFSSQVCRVAYEIMQ 480

DB 421 YSNYEFISDINSISWAGLHNRNTENSLRGVILDIHFLSQADPLVCTFSSQVCRVAYEIMQ 480

QY 481 TLHPDASANFSLDDIYFVGQNAHNOIAVYPHOPRTKEEIPMEPGDIIGVAGNHNGYS 540

DB 481 TLHPDASANFSLDDIYFVGQNAHNOIAVYPHOPRTKEEIPMEPGDIIGVAGNHNGYS 540

QY 541 KGVNRKLGKTGLYPSYKVKREKIEIVKYPTYPEAEK 575

DB 541 KGVNRKLGKTGLYPSYKVKREKIEIVKYPTYPEAEK 575

## RESULT 4

ABB08405

ID ABB08405 standard; Protein; 575 AA.

XX AC ABB08405;

XX DT 07-MAY-2002 (first entry)

XX DE Alpha1,6-fucosyl transferase amino acid sequence.

XX KW Plant; glycoprotein; alpha1,6-fucosyl transferase; alpha1,6-Ft;

XX KW enzyme.

XX OS Homo sapiens.

XX PN JP2001333787-A.

XX PD 04-DEC-2001.

XX PF 06-MAR-2001; 2001JP-0063704.

XX XX



PR 22-MAR-2000; 2000JP-0081059.  
 XX (TANI/) TANIGUCHI N.  
 PA (SEKI/) SEKI T.  
 XX (FUJI/) FUJIYAMA K.  
 XX  
 DR WPI; 2002-158816/21.  
 DR N-PSDB; ABA98809.  
 XX  
 PT A plant cell with an animal type sugar chain adding function, for the  
 PT preparation of a glycoprotein with an animal type sugar chain -  
 XX  
 PS Example 1; Page 24-26; 38pp; Japanese.  
 XX  
 CC The invention relates to a plant cell with an animal type sugar chain  
 CC adding function, created by transforming a tobacco cell with a gene  
 CC encoding an enzyme derived from an animal which can transfer a fucose  
 CC residue to the reductive end acetylglucosamine residue of a sugar chain.  
 CC The gene that is introduced into the plant cell encodes the enzyme  
 CC alpha1,6-fucosyl transferase. The method of the invention is useful for  
 CC the preparation of a glycoprotein having animal type sugar chain. The  
 CC current sequence represents alpha1,6-fucosyl transferase amino acid  
 CC sequence.  
 XX  
 SQ Sequence 575 AA;  
 Query Match 96.9%; Score 2984; DB 23; Length 575;  
 Best Local Similarity 96.2%; Pred. No. 3.9e-263;  
 Matches 553; Conservative 10; Mismatches 12; Indels 0; Gaps 0;  
 QY 1 MRWGTGSRWIMLILFWAGTLLFYGGHLVRNDHPDHSRELSKILAKLERLKKQNEDL 60  
 DB 1 MRWGTGSRWIMLILFWAGTLLFYGGHLVRNDHPDHSRELSKILAKLERLKKQNEDL 60  
 QY 61 RMAESLRIPGPIDOGTATGRVRLVLEOLVKAKEQIENYKKAQANDLKGKHEILRRRIE 120  
 DB 61 RMAESLRIPGPIDOGTATGRVRLVLEOLVKAKEQIENYKKAQANDLKGKHEILRRRIE 120  
 QY 121 NGAKELWFFLQSELKKLKEGNEQLQRADEILDLGHHERSIMTDLVYLSQTDGAGDWR 180  
 DB 121 NGAKELWFFLQSELKKLKEGNEQLQRADEILDLGHHERSIMTDLVYLSQTDGAGDWR 180  
 QY 181 EKEAKDLTBLVORRTYLPNDKCSKARKLVNINKGCGYQGLHHVYVCFMAYGTORT 240  
 DB 181 EKEAKDLTBLVORRTYLPNDKCSKARKLVNINKGCGYQGLHHVYVCFMAYGTORT 240  
 QY 241 LILESQNWRYATGGWETVPRPVSCTDRSGLSGTHWSGEVKDKNVQVVELPIVDSLHPR 300  
 DB 241 LILESQNWRYATGGWETVPRPVSCTDRSGLSGTHWSGEVKDKNVQVVELPIVDSLHPR 300  
 QY 301 PPYLPLAVPEDLADRLRLRHGDPVWVWSQFVKYLIRPQWLERIEIETTKLGFKHPVI 360  
 DB 301 PPYLPLAVPEDLADRLRLRHGDPVWVWSQFVKYLIRPQWLERIEIETTKLGFKHPVI 360  
 QY 361 GVHVRTDKVGEAAPHPIEYVNVHVEHFOLLERMKVKKRVYLATDDPSLLKEATK 420  
 DB 361 GVHVRTDKVGEAAPHPIEYVNVHVEHFOLLERMKVKKRVYLATDDPSLLKEATK 420  
 QY 421 YSNYFISDNTSWAGLHNRVTENSLRGVILDIHFLSQADFLVCTFSQCVRAVEIMQ 480  
 DB 421 YSNYFISDNTSWAGLHNRVTENSLRGVILDIHFLSQADFLVCTFSQCVRAVEIMQ 480  
 QY 481 TLHPDASANFHSLLDIYFGGQNAJNIAVYHPQRTKEETPMPEGDIIGVAGNHNGYS 540  
 DB 481 TLHPDASANFHSLLDIYFGGQNAJNIAVYHPQRTKEETPMPEGDIIGVAGNHNGYS 540  
 QY 541 KGNVRKLGRTGLYPSYKVKREKTIETVKYPTYPEAK 575  
 DB 541 KGNVRKLGRTGLYPSYKVKREKTIETVKYPTYPEAK 575

RESULT 5  
 AAW22124

ID AAW22124 standard; Protein; 575 AA.  
 XX AC AAW22124;  
 XX DT 05-MAR-1998 (first entry)  
 XX DE Pig alpha 1-6 fucosyltransferase.  
 XX KW Alpha 1-6 fucosyltransferase; enzyme; pig; human; fucose transfer;  
 KW guanosine diphosphate; sugar chain synthesis; modification; antibody;  
 KW GlcNAc; cancer diagnosis.  
 XX OS Sus scrofa.  
 XX PN W09727303-A1.  
 XX PD 31-JUL-1997.  
 XX PF 23-JAN-1997; 97WO-JP00171.  
 XX PR 22-JUL-1996; 96JP-0192260.  
 PR 24-JAN-1996; 96JP-0010365.  
 PR 21-JUN-1996; 96JP-0161648.  
 PR 24-JUN-1996; 96JP-0162813.  
 XX (TOYM) TOYO BOSEKI KK.  
 PA Shiba T, Taniguchi N, Uozumi N, Yanagidani S;  
 PI WPI; 1997-393690/36.  
 XX N-PSDB; AAT76573.  
 DR Human or pig alpha 1-6 fucosyltransferase and DNA encoding it - for  
 PT synthesis and modification of sugar chains and used as an antigen  
 PT for production of diagnostic antibodies  
 XX Claim 4; Page 30-34; 61pp; Japanese.  
 CC AAW22124 and AAW22125 represent the pig and human alpha 1-6  
 CC fucosyltransferases of the invention, respectively. The enzyme transfers  
 CC fucose from guanosine diphosphate to the 6-hydroxyl group of the GlcNAc  
 CC nearest to R in the receptor molecule: (GlcNAc-beta 1-2Man-alpha 1-6)  
 CC (GlcNAc-beta 1-2Man-alpha 1-3)Man-beta 1-4GlcNAc-beta 1-4GlcNAc-R to give  
 CC (GlcNAc-beta 1-2Man-alpha 1-6)(GlcNAc-beta 1-2Man-alpha 1-3)Man-beta  
 CC 1-4GlcNAc-beta 1-4(Fuc-alpha 1-6)GlcNAc-R. It has an optimum pH of about  
 CC 7.0 (pig) or 7.5 (human), and is stable over the pH range 4-10 after 5  
 CC hours at 4 degrees C. The optimum working temperature of the  
 CC alpha 1-6 fucosyltransferases is 30-37 degrees C. A bivalent metal is  
 CC not required for activity of the enzyme, and the enzyme is not inhibited  
 CC in the presence of 5 mM EDTA. The enzyme is useful in the synthesis and  
 CC modification of sugar chains, and as antigen for the production of  
 CC antibodies recognising the enzyme. The antibodies can be used for the  
 CC diagnosis of cancer and other diseases.  
 XX SQ Sequence 575 AA;  
 Query Match 93.6%; Score 2885; DB 18; Length 575;  
 Best Local Similarity 92.9%; Pred. No. 4.3e-254;  
 Matches 534; Conservative 16; Mismatches 25; Indels 0; Gaps 0;  
 QY 1 MRWGTGSRWIMLILFWAGTLLFYGGHLVRNDHPDHSRELSKILAKLERLKKQNEDL 60  
 DB 1 MRWGTGSRWIMLILFWAGTLLFYGGHLVRNDHPDHSRELSKILAKLERLKKQNEDL 60  
 QY 61 RMAESLRIPGPIDOGTATGRVRLVLEOLVKAKEQIENYKKAQANDLKGKHEILRRRIE 120  
 DB 61 RMAESLRIPGPIDOGTATGRVRLVLEOLVKAKEQIENYKKAQANDLKGKHEILRRRIE 120  
 QY 121 NGAKELWFFLQSELKKLKEGNEQLQRADEILDLGHHERSIMTDLVYLSQTDGAGDWR 180  
 DB 121 NGAKELWFFLQSELKKLKEGNEQLQRADEILDLGHHERSIMTDLVYLSQTDGAGDWR 180  
 QY 181 EKEAKDLTBLVORRTYLPNDKCSKARKLVNINKGCGYQGLHHVYVCFMAYGTORT 240

Db 181 EKEARDLTQLVQRRTTYLQNPDKCSKAKKLVNINKGCGYGCQLHHVVCYCFMAYGTQRT 240  
Qy 241 LILESQWRYATGCGWETVFRPVSECTDRSLGSLTGHWSGEVKKVNOVVELPIVDSLHPR 300  
Db 241 LALESHNNRYATGCGWETVFRPVSECTDRSGSSTGHWSGEVKKVNOVVELPIVDSLHPR 300  
Qy 301 PPYLPLAVPEDLADRLVHVGDPAVWVVSQFVKYLIRPQWLERIEETTKKLGFKHPVI 360  
Db 301 PPYLPLAVPEDLADRLVHVGDPAVWVVSQFVKYLIRPQWLERIEETTKKLGFKHPVI 360  
Qy 361 GVHVRTDKVTEAAFPDIEYMWVHEHFPOLLERMMKVDKRVYLAATDDPSLLKEAKTK 420  
Db 361 GVHVRTDKVTEAAFPDIEYMWVHEHFPOLLERMMKVDKRVYLAATDDPSLLKEAKTK 420  
Qy 421 YSNVEFISDNTSISWSAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSQVCRVAYEIMQ 480  
Db 421 YPSVEFISDNTSISWSAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSQVCRVAYEIMQ 480  
Qy 481 TLHPDASANFSLDDIYYFGGQNAHNOIAVYPHQPRTKEEIPMEPGDIIGVAGNHWNGYS 540  
Db 481 ALHPDASANFSLDDIYYFGGQNAHNOIAVYPHQPRTKEEIPMEPGDIIGVAGNHWNGYS 540  
Qy 541 KGVNRKLGKTGLPSYKVKREKIEYKTYPTYPEAEK 575  
Db 541 KGVNRKLGKTGLPSYKVKREKIEYKTYPTYPEADK 575

## RESULT 6

AAG73884  
ID AAG73884 standard; Protein; 515 AA.  
XX  
AC AAG73884;  
XX  
DT 03-SEP-2001 (first entry)  
XX  
DE Human colon cancer antigen protein SEQ ID NO:4648.  
XX  
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KW colorectal carcinoma; chromosome 14.  
XX  
OS Homo sapiens.  
XX  
PN WO200122920-A2.  
XX  
PD 05-APR-2001.  
XX  
PF 28-SEP-2000; 2000WO-US26524.  
XX  
PR 29-SEP-1999; 99US-0157137.  
XX  
PR 03-NOV-1999; 99US-0163280.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SM, Barash SC, Birse CE, Rosen CA;  
XX  
PI WPI; 2001-235357/24.  
XX  
DR N-P5DB; AAH33315.  
XX

Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
PT useful for preventing, diagnosing and/or treating colorectal cancers -  
XX  
PS Claim 11; Page 6451-6453; 9803pp; English.  
XX  
AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
CC the proteins are collectively known as colon cancer antigens. The colon  
CC cancer antigens have cytostatic activity and can be used in gene  
CC therapy and vaccine production. N and P may be used in the prevention,  
CC diagnosis and treatment of diseases associated with inappropriate P  
CC expression. For example, N and P may be used to treat disorders  
CC associated with decreased expression by rectifying mutations or deletions  
CC in a patient's genome that affect the activity of P by expressing

CC inactive proteins or to supplement the patients own production of P.  
CC Additionally, N may be used to produce the colon cancer-associated Ps,  
CC by inserting the nucleic acids into a host cell and culturing the cell  
CC to express the proteins. N and P can be used in the prevention, diagnosis  
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
CC and AAG77789 represent sequences used in the exemplification of the  
CC present invention.  
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
CC missing at time of publication, meaning no sequences are present for  
CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
XX

## SQ Sequence 515 AA;

Query Match 85.7%; Score 2639; DB 22; Length 515;  
Best Local Similarity 95.7%; Pred. No. 1e-231;  
Matches 489; Conservative 11; Mismatches 11; Indels 0; Gaps 0;  
Qy 65 ESLRIPGPDIDGQATGRVRLERQVLKAKEIQENYKQARNDLGKHIEILRRRIENGAK 124  
Db 5 KSLRIPGPDIDGQPAJGRVRLERQVLKAKEIQENYKQARNDLGKHIEILRRRIENGAK 64  
Qy 125 ELWFLOSELKLLKLSGNEIQRHADEILDLGHHSIMTDLYVLSQDAGDWREKEA 184  
Db 65 ELWFLOSELKLLKLSGNEIQRHADEFLDLGHHSIMTDLYVLSQDAGDWREKEA 124  
Qy 185 KDLTELQRRITTYLQNPDKCSKAKKLVNINKGCGYGCQLHHVVCYCFMAYGTQRTLILE 244  
Db 125 KDLTELQRRITTYLQNPDKCSKAKKLVNINKGCGYGCQLHHVVCYCFMAYGTQRTLILE 184  
Qy 245 SQNRYATGCGWETVFRPVSECTDRSLGSLTGHWSGEVKKVNOVVELPIVDSLHPRPPL 304  
Db 185 SQNRYATGCGWETVFRPVSECTDRSGTSTGHWSGEVKKVNOVVELPIVDSLHPRPPL 244  
Qy 305 PLAYPEDLADRLVHVGDPAVWVVSQFVKYLIRPQWLERIEETTKKLGFKHPVIGVHV 364  
Db 245 PLAYPEDLADRLVHVGDPAVWVVSQFVKYLIRPQWLERIEETTKKLGFKHPVIGVHV 304  
Qy 365 RRTDKVGTAAAFHPIEYMWVHEHFPOLLERMMKVDKRVYLAATDDPSLLKEAKTKYSNY 424  
Db 305 RRTDKVGTAAAFHPIEYMWVHEHFPOLLARMQVKKRVYLAATDDPSLLKEAKTKYNY 364  
Qy 425 EFISDNTSISWSAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSQVCRVAYEIMQTLHP 484  
Db 365 EFISDNTSISWSAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSQVCRVAYEIMQTLHP 424  
Qy 485 DASANFHSLLDDIYYFGGQNAHNOIAVYPHQPRTKEEIPMEPGDIIGVAGNHWNGYSKGVN 544  
Db 425 DASANFHSLLDDIYYFGGQNAHNOIAVYAHQPRTADEIPMEPGDIIGVAGNHWNGYSKGVN 484  
Qy 545 RKLKGTGLPSYKVKREKIEYKTYPTYPEAEK 575  
Db 485 RKLKGTGLPSYKVKREKIEYKTYPTYPEAEK 515

## RESULT 7

AAH75061  
ID AAB75061 standard; Protein; 339 AA.  
XX

## AC AAB75061;

XX  
DT 20-JUL-2001 (first entry)  
XX

DE Human alpha 1-6 fucosyltransferase protein 237-575 SEQ ID NO:3.  
XX

KW Human; alpha 1-6 fucosyltransferase; alpha 1-6 fucT; antibody;  
KW alpha 1-6 fucose transferase; anti-human alpha 1-6 fucT; immunoassay.  
XX

OS Homo sapiens.  
XX

PN JP2001011097-A.  
XX

PD 16-JAN-2001.  
XX

```

PF 29-JUN-1999; 99JP-0183569.
XX
PR 29-JUN-1999; 99JP-0183569.
XX
PA (FJURE ) FUJIREBIO KK.
XX
DR WPI: 2001-275926/29.
DR N-PSDB; AAF87952.
XX
PT Novel anti-human alpha1-6 fucose transferase antibody useful for
PT immunoassay -
XX
PS Example 2; Page 8-9; 11pp; Japanese.
XX
CC The present invention describes an anti-human alpha 1-6
CC fucosyltransferase (alpha 1-6 fuct, also called alpha 1-6 fucose
CC transferase) antibody. Also described are: (1) a hybridoma producing
CC the above monoclonal antibody; (2) an immunoassay for detecting human
CC alpha 1-6 fuct by using the above antibody or its antibody fragment;
CC and (3) a reagent used for the above immunoassay. The anti-human
CC alpha 1-6 fuct antibody can be used for immunoassay. The present
CC sequence represents the human alpha 1-6 fuct protein of residues 237
CC to 575, which is used in an example from the present invention.
XX
SQ Sequence 339 AA;
Query Match 57.5%; Score 1771; DB 22; Length 339;
Best Local Similarity 96.2%; Pred. No. 8.9e-153;
Matches 326; Conservative 8; Mismatches 5; Indels 0; Gaps 0;
QY 237 TQRTLLSQNRWYATGCGWETVFRPVSETCTDRSLGTHSGEVKDKNVQVVELPIYDS 296
DB 1 TQRTLLSQNRWYATGCGWETVFRPVSETCTDRSLGTHSGEVKDKNVQVVELPIYDS 60
QY 297 LHPRPPYPLAVPEDLADRLRVHGDPAVWWSQFVKYLIRPOPWLEIEETTKLGFK 356
DB 61 LHPRPPYPLAVPEDLADRLRVHGDPAVWWSQFVKYLIRPOPWLEIEETTKLGFK 120
QY 357 HPVIGVHVVRTDKVGTAAAFHPIEYVMVHVEHFQLLERMKVDKRVYLATDDPSLLKE 416
DB 121 HPVIGVHVVRTDKVGTAAAFHPIEYVMVHVEHFQLLERMKVDKRVYLATDDPSLLKE 180
QY 417 AKTKYSNVEFISDINSISAGLHNYRTENS LGVILDIHFLSQADFLVCTFSSQVCRVAY 476
DB 181 AKTKYPNVEFISDINSISAGLHNYRTENS LGVILDIHFLSQADFLVCTFSSQVCRVAY 240
QY 477 EIMQTLHPDASANFHSLODIYFVGQNAHQIAVYPHOPRTKEEIPMEPGDIIIGVAGNHW 536
DB 241 EIMQTLHPDASANFHSLODIYFVGQNAHQIAVYPHOPRTKEEIPMEPGDIIIGVAGNHW 300
QY 537 NGYSGVNRKLGKTGLYPSYKVKREKIEYKTYPTYPEAEK 575
DB 301 DQYSGVNRKLGKTGLYPSYKVKREKIEYKTYPTYPEAEK 339
RESULT 8
ABB59117
ID ABB59117 standard; Protein; 619 AA.
XX
AC ABB59117;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 4143.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
XX 27-SEP-2001.
XX
23-MAR-2001; 2001WO-US09231.
XX
23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI: 2001-656860/75.
DR N-PSDB; ABL03220.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 4143; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 619 AA;
Query Match 48.3%; Score 1489.5; DB 22; Length 619;
Best Local Similarity 46.6%; Pred. No. 1.1e-126;
Matches 285; Conservative 110; Mismatches 168; Indels 49; Gaps 7;
QY 3 AWTGSW-RWIMLILFANGTLLFYIGCHLVDRN-----DHPDSSRELSKILAKLERLKKQ 56
DB 10 ASANSWARALIFLWAGLVYFVVKLTNTQGOAAGESELNARRISQALOMLEHTQR 69
QY 57 NEDLRRMAESLRIPEGPIDOQTATGRVRLVEEQVLVKAKEQIE----- 98
DB 70 NEELQQLIDELMSQ--LDKOSAMKLVQRLNDALNPKLAPEVAGPEPESFESAPADLR 127
QY 99 ---NYKQARNDL-----GKDHEILRRRIENGAKELMFFQSELSKUKKL--- 140
DB 128 GWNVAEGAPNDLEAGVPDHGCEPEPSLEYEFTRRRIQTNICEIWNFFSELGKVKAVAA 187
QY 141 --EGNELQORHADEILLDLGHHERSIMTDLYLVSQTDGAGWEKREKADLTQLVQRITYL 198
DB 188 GHASADLEESINQVLLQGAEHKRLSLSDMERMRQSDGYEARWHRKEARDLSLVQRLHL 247
QY 199 QNPDKCKARKLVNINKGCGYGCQLHHVYCFMAYGTQRTLLILESQNWRYATGGWETV 258
DB 248 QNPSCQNAKRLVCKLNGKCGYGCQLHHVYCFIVAVATERTLLKSGRWYHKGWEEV 307
QY 259 FRPVSETCTDRSLGTHSGEVKDKNVQVVELPIVDSLHPRPPYPLAVPEDLADRLLR 318
DB 308 FQPVSNCHDAGTANTYNWPK---PNTQVLVLPFIIDSLMPRPYPLAVPEDLAPRLKR 364
QY 319 VHGDPVWWSQFVKYLIRPOPWLEIEETTKLGFKHPVIGVHVVRTDKVGTAAAFH 378
DB 365 LHGDFIVWVGQFLKYLRLPQPTTRDFLTSGMRNLGWERPIVGVHVVRTDKVGTAAACHS 424
QY 379 IEEYVMVHVEHFQLLERMKVDKRVYLATDDPSLLKEAKTKYSNVEFISDINSISAGL 438
DB 425 VEEYTWTVEDYRTLEVNGSTVARRIFLASDAAQVIEARRKYQYQYIIGPEVARMASV 484
QY 439 HNYRTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQTLHPDASANFHSLODIY 498
DB 485 STRVTD TALNGIILDIHLLMSDHLVCTFSSQVCRVAYEIMQTLHPDAAHFKSLDDIY 544

```



```
ABB28797
ID ABB28797 standard; Peptide; 82 AA.
XX
AC ABB28797;
XX
DT 01-FEB-2002 (first entry)
XX
DE Peptide #1448 encoded by breast cell single exon nucleic acid probe.
XX
KW Human; microarray; single exon probe; gene expression; breast;
XX
KW disease; cancer.
XX
XX Homo sapiens.
OS
XX WO200157271-A2.
PN
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00662.
PF
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
XX Claim 27; SEQ ID NO 11765; 327pp + sequence listing; English.
PS
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic nucleic acid
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 82 AA;
SQ
Query Match 14.0%; Score 431; DB 22; Length 82;
Best Local Similarity 96.3%; Pred. No. 2e-31;
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 280 EVKDKNQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 339
DB 1 EVKDKNQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 60

QY 340 PWLEREIEETTKLGFKHPVIG 361
DB 61 PWLEREIEETTKLGFKHPVIG 82

RESULT 13
ABB19422
ID ABB19422 standard; Protein; 82 AA.
XX
XX ABB19422;
AC
XX 23-JAN-2002 (first entry)
DT
```

```
RESULT 12
ABB33985
ID ABB33985 standard; Peptide; 82 AA.
XX
XX ABB33985;
AC
XX 04-FEB-2002 (first entry)
DT
XX Peptide #1491 encoded by human foetal liver single exon probe.
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
XX Homo sapiens.
OS
XX WO200157277-A2.
PN
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00669.
PF
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
XX Claim 27; SEQ ID NO 26620; 639pp + sequence listing; English.
PS
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 82 AA;
SQ
Query Match 14.0%; Score 431; DB 22; Length 82;
Best Local Similarity 96.3%; Pred. No. 2e-31;
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 280 EVKDKNQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 339
DB 1 EVKDKNQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 60

QY 340 PWLEREIEETTKLGFKHPVIG 361
DB 61 PWLEREIEETTKLGFKHPVIG 82

RESULT 13
ABB19422
ID ABB19422 standard; Protein; 82 AA.
XX
XX ABB19422;
AC
XX 23-JAN-2002 (first entry)
DT
```

XX DE Protein #1421 encoded by probe for measuring heart cell gene expression.  
XX KW Human; gene expression; heart; microarray; vascular system;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
XX congenital heart disease.  
XX OS Homo sapiens.  
XX PN WO200157274-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US00666.  
XX PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488899/53.  
XX DR Single exon nucleic acid probes for analyzing gene expression in human hearts -  
XX PT hearts -  
XX PT hearts -  
XX PS Claim 15; SEQ ID NO 21192; 530pp; English.  
XX CC The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart (see ABA21535-ABA41305). The present sequence is a protein encoded by one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease.  
XX CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 82 AA;  
Query Match 14.0%; Score 431; DB 22; Length 82;  
Best Local Similarity 96.3%; Pred. No. 2e-31;  
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Qy 280 EVKDKNVQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 339  
Db 1 EVKDKNVQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 60  
Qy 340 PWLEREIEETTKLGFKHPVIG 361  
Db 61 PWLEREIEETTKLGFKHPVIG 82  
RESULT 14  
AAM54747  
ID AAM54747 standard; Protein; 82 AA.  
XX AC AAM54747;  
XX DT 05-NOV-2001 (first entry)  
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 26852.  
XX KW Human; brain expressed exon; gene expression analysis; probe;

KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.  
XX KW Homo sapiens.  
XX PN WO200157275-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US00667.  
XX PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483446/52.  
XX DR Single exon nucleic acid probes for analyzing gene expression in human brains -  
XX PT brains -  
XX PS Example 4; SEQ ID NO: 26852; 650pp + Sequence Listing; English.  
XX CC The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is a protein encoded by one of the probes of the invention.  
XX SQ Sequence 82 AA;  
Query Match 14.0%; Score 431; DB 22; Length 82;  
Best Local Similarity 96.3%; Pred. No. 2e-31;  
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Qy 280 EVKDKNVQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 339  
Db 1 EVKDKNVQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 60  
Qy 340 PWLEREIEETTKLGFKHPVIG 361  
Db 61 PWLEREIEETTKLGFKHPVIG 82  
RESULT 15  
AAM67139  
ID AAM67139 standard; Protein; 82 AA.  
XX AC AAM67139;  
XX DT 06-NOV-2001 (first entry)  
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 27445.  
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma.  
XX OS Homo sapiens.  
XX PN WO200157276-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US00668.

```

XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 27445; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention.
XX
XX Sequence 82 AA;
SQ
Query Match 14.0%; Score 431; DB 22; Length 82;
Best Local Similarity 96.3%; Pred. No. 2e-31;
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 280 EVKDKNVQVVELPIVDSLHPRPPYLPLAVPEDLADRLRVHGDPAVWVSQFVKYLIRPQ 339
Db 1 EVKDKNVQVVELPIVDSLHPRPPYLPLAVPEDLADRLRVHGDPAVWVSQFVKYLIRPQ 60
QY 340 PWLEREIEETTKKLGFKHPVIG 361
Db 61 PWLEREIEETTKKLGFKHPVIG 82

```

Search completed: February 2, 2004, 08:41:46  
Job time : 40.5 secs





```
; MOLECULE TYPE: peptide
US-08-913-805A-10

Query Match
Best Local Similarity 96.2%; Score 2984; DB 3; Length 575;
Matches 553; Conservative 10; Mismatches 12; Indels 0; Gaps 0;

QY 1 MRWTGSRWIMLILFAWGTLFLFYGGHLVRNDHPDHSRELSKILAKLERLKKQONEDL 60
DB 1 MRPTGSRWIMLILFAWGTLFLFYGGHLVRNDHPDHSRELSKILAKLERLKKQONEDL 60

QY 61 RMAESLRIPGPIQOGTATGRVRLVEQLVKAKEIQENYKQANDLKGKDEILRRRIE 120
DB 61 RMAESLRIPGPIQOGTATGRVRLVEQLVKAKEIQENYKQANDLKGKDEILRRRIE 120

QY 121 NGAKELWFFLQSELKLLKLEGNELQRADEFLDLGHHSRIMTDLVYLSQTDGAGWR 180
DB 121 NGAKELWFFLQSELKLLKLEGNELQRADEFLDLGHHSRIMTDLVYLSQTDGAGWR 180

QY 181 EKEADLTTELVRRTITLQNPDKCSKARKLVNINKGCGYQGLHHVVCYCFMAYGTQRT 240
DB 181 EKEADLTTELVRRTITLQNPDKCSKARKLVNINKGCGYQGLHHVVCYCFMAYGTQRT 240

QY 241 LILESQNRWYATGCGWETVFRPVSETCTDRSGISTGHSGEVKDKNVQVVELPIVDSLHPR 300
DB 241 LILESQNRWYATGCGWETVFRPVSETCTDRSGISTGHSGEVKDKNVQVVELPIVDSLHPR 300

QY 301 PPYLPLAVPEDLADRLRVRHGDPAVWVSQFVKYLIRPOPMLEREIEETTKLGFKHPVI 360
DB 301 PPYLPLAVPEDLADRLRVRHGDPAVWVSQFVKYLIRPOPMLEREIEETTKLGFKHPVI 360

QY 361 GVHVRTDKVGTAAAFPIEYVHVEHFQLLARRMQVKKRVYLATDDPSLLKEATK 420
DB 361 GVHVRTDKVGTAAAFPIEYVHVEHFQLLARRMQVKKRVYLATDDPSLLKEATK 420

QY 421 YSNYEFISDNSISWSAGLHNRYTENSRLRGVILDIHFLSQADFLVCTFSSQVCVAYEIMQ 480
DB 421 YSNYEFISDNSISWSAGLHNRYTENSRLRGVILDIHFLSQADFLVCTFSSQVCVAYEIMQ 480

QY 481 TLHPDASANFSLDDIYFPGQNAHQIAVYPHQRTKEEIPMEPGDIIGVAGNHNWGS 540
DB 481 TLHPDASANFSLDDIYFPGQNAHQIAVYPHQRTKEEIPMEPGDIIGVAGNHNWGS 540

QY 541 KGNVRKLGKGLYPSYKVRKIEIVKYPTYPEAEK 575
DB 541 KGNVRKLGKGLYPSYKVRKIEIVKYPTYPEAEK 575

RESULT 2
US-09-442-629-10
; Sequence 10, Application US/09442629
; Patent No. 6291219
; GENERAL INFORMATION:
; APPLICANT: TANIGUCHI, Naoyuki
; UOZUMI, Naofumi
; SHIBA, Tetsuo
; YANAGIDANI, Shusaku
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: US
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3+ Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
; SOFTWARE: WordPerfect 6.1 Windows
; CURRENT APPLICATION DATA: US/09/442,629
; APPLICATION NUMBER: US/09/442,629

; FILING DATE: 18-No. 6291219-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/913.805A
; FILING DATE: 7 JAN 1998
; APPLICATION NUMBER: PCT/JP97/00171
; FILING DATE: 23 JAN 1997
; APPLICATION NUMBER: JP 192260
; FILING DATE: 22 JUL 1996
; APPLICATION NUMBER: JP 162813
; FILING DATE: 24 JUN 1996
; APPLICATION NUMBER: JP 161648
; FILING DATE: 21 JUN 1996
; APPLICATION NUMBER: JP 10365
; FILING DATE: 24 JAN 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Toffenetti, Judith L.
; REGISTRATION NUMBER: 39,048
; REFERENCE/DOCKET NUMBER: 2356/3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-429-1776
; TELEFAX: 202-429-0796
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 575 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-442-629-10

Query Match 96.9%; Score 2984; DB 3; Length 575;
Best Local Similarity 96.2%; Pred. No. 6.1e-270;
Matches 553; Conservative 10; Mismatches 12; Indels 0; Gaps 0;

QY 1 MRWTGSRWIMLILFAWGTLFLFYGGHLVRNDHPDHSRELSKILAKLERLKKQONEDL 60
DB 1 MRPTGSRWIMLILFAWGTLFLFYGGHLVRNDHPDHSRELSKILAKLERLKKQONEDL 60

QY 61 RMAESLRIPGPIQOGTATGRVRLVEQLVKAKEIQENYKQANDLKGKDEILRRRIE 120
DB 61 RMAESLRIPGPIQOGTATGRVRLVEQLVKAKEIQENYKQANDLKGKDEILRRRIE 120

QY 121 NGAKELWFFLQSELKLLKLEGNELQRADEFLDLGHHSRIMTDLVYLSQTDGAGWR 180
DB 121 NGAKELWFFLQSELKLLKLEGNELQRADEFLDLGHHSRIMTDLVYLSQTDGAGWR 180

QY 181 EKEADLTTELVRRTITLQNPDKCSKARKLVNINKGCGYQGLHHVVCYCFMAYGTQRT 240
DB 181 EKEADLTTELVRRTITLQNPDKCSKARKLVNINKGCGYQGLHHVVCYCFMAYGTQRT 240

QY 241 LILESQNRWYATGCGWETVFRPVSETCTDRSGISTGHSGEVKDKNVQVVELPIVDSLHPR 300
DB 241 LILESQNRWYATGCGWETVFRPVSETCTDRSGISTGHSGEVKDKNVQVVELPIVDSLHPR 300

QY 301 PPYLPLAVPEDLADRLRVRHGDPAVWVSQFVKYLIRPOPMLEREIEETTKLGFKHPVI 360
DB 301 PPYLPLAVPEDLADRLRVRHGDPAVWVSQFVKYLIRPOPMLEREIEETTKLGFKHPVI 360

QY 361 GVHVRTDKVGTAAAFPIEYVHVEHFQLLARRMQVKKRVYLATDDPSLLKEATK 420
DB 361 GVHVRTDKVGTAAAFPIEYVHVEHFQLLARRMQVKKRVYLATDDPSLLKEATK 420

QY 421 YSNYEFISDNSISWSAGLHNRYTENSRLRGVILDIHFLSQADFLVCTFSSQVCVAYEIMQ 480
DB 421 YSNYEFISDNSISWSAGLHNRYTENSRLRGVILDIHFLSQADFLVCTFSSQVCVAYEIMQ 480

QY 481 TLHPDASANFSLDDIYFPGQNAHQIAVYPHQRTKEEIPMEPGDIIGVAGNHNWGS 540
DB 481 TLHPDASANFSLDDIYFPGQNAHQIAVYPHQRTKEEIPMEPGDIIGVAGNHNWGS 540

QY 541 KGNVRKLGKGLYPSYKVRKIEIVKYPTYPEAEK 575
DB 541 KGNVRKLGKGLYPSYKVRKIEIVKYPTYPEAEK 575
```

## RESULT 3

US-08-913-805A-2  
 ; Sequence 2, Application US/08913805A  
 ; Patent No. 6054304  
 ; GENERAL INFORMATION:  
 ; APPLICANT: TANIGUCHI, Naoyuki  
 ; APPLICANT: UOZUMI, Naofumi  
 ; APPLICANT: SHIBA, Tetsuo  
 ; APPLICANT: YANAGIDANI, Shusaku  
 ; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
 ; NUMBER OF SEQUENCES: 15  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Kenyon & Kenyon  
 ; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
 ; CITY: Washington  
 ; STATE: DC  
 ; COUNTRY: US  
 ; ZIP: 20036  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3+ Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
 ; SOFTWARE: WordPerfect 6.1 Windows  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/913.805A  
 ; FILING DATE: 7 JAN 1998  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/JP97/00171  
 ; FILING DATE: 23 JAN 1997  
 ; APPLICATION NUMBER: JP 192260  
 ; FILING DATE: 22 JUL 1996  
 ; APPLICATION NUMBER: JP 162813  
 ; FILING DATE: 24 JUN 1996  
 ; APPLICATION NUMBER: JP 161648  
 ; FILING DATE: 21 JUN 1996  
 ; APPLICATION NUMBER: JP 10365  
 ; FILING DATE: 24 JAN 1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Toffenetti, Judith L.  
 ; REGISTRATION NUMBER: 39,048  
 ; REFERENCE/DOCKET NUMBER: 2356/3  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 202-429-1776  
 ; TELEFAX: 202-429-0796  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 575 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-913-805A-2

Query Match 93.6%; Score 2885; DB 3; Length 575;  
 Best Local Similarity 92.9%; Pred. No. 1.1e-260;  
 Matches 534; Conservative 16; Mismatches 25; Indels 0; Gaps 0;  
 QY 1 MRWTGSRWIMLILFANGTLLFYIGGHLVRDNDHPDHSRSLSKILAKLERLKQONEDL 60  
 DB 1 MRPTGSRWIMLILFANGTLLFYIGGHLVRDNDHSDHSRSLSKILAKLERLKQONEDL 60  
 QY 61 RMAESLRIPGPIDQGTATGRVRLBEQLVAKAEQIENYKKQARNDLGKDEILRRRIE 120  
 DB 61 RMAESLRIPGPIDQGPASGRVRLAEQFMKAEQIENYKKQTKNGPGKDEILRRRIE 120  
 QY 121 NGAKELWFLQSELKKLKEGNELOLQHADELILLDGHHSRIMTDLVYLSQTDGAGWR 180  
 DB 121 NGAKELWFLQSELKKLKEGNELOLQHADELFLSLDGHHSRIMTDLVYLSQTDGAGWR 180  
 QY 181 EKEAKDLTELVRRIITYLQNPDKCSKAKLVNCKGCGYCOLHHVVCYCFMIAYGRTQRT 240

DB 181 EKEAKDLTELVRRIITYLQNPDKCSKAKLVNCKGCGYCOLHHVVCYCFMIAYGRTQRT 240  
 QY 241 LILESQNRVYATGGWETVFRPVSCTCTDRSGLSTGHSWGEVKQKNVQVVELPIVDSLHPR 300  
 DB 241 LALESHNRVYATGGWETVFRPVSCTCTDRSGSSTGHSWGEVKQKNVQVVELPIVDSVHPR 300  
 QY 301 PPYLPLAVPEDLADRLRLRVHGDPAVWVWSQFVKYLIRPQWLEREIBETTKLGFKHPVI 360  
 DB 301 PPYLPLAVPEDLADRLVRVHGDPAVWVWSQFVKYLIRPQWLEREIBETTKLGFKHPVI 360  
 QY 361 GVHVRRTDKVGTEAAHPHIEEYVHVVEEHFQLLERRMKVKRYVYLATDDPSLLKEATK 420  
 DB 361 GVHVRRTDKVGAAEAFPHIEEYTVHVEEDFQLLARRMQVKRYVYLATDDPSLLKEATK 420  
 QY 421 YSNYEFTSDNSISWSAGLHNRYTENSRLGVILDIHFLSQADFLVCTFSQVCRVAYEIMQ 480  
 DB 421 YPSYEFTSDNSISWSAGLHNRYTENSRLGVILDIHFLSQADFLVCTFSQVCRVAYEIMQ 480  
 QY 481 TLHPDASANFHSLDIYYFGQNAHNOIAYVPHQRTKEIIPMPGDIIGVAGNHNWGS 540  
 DB 481 ALHPDASANFRSLDIYYFGQNAHNOIAYVPHQRTKEIIPMPGDIIGVAGNHNWGS 540  
 QY 541 KGVNRKLGKTGLYPSYKVKREKIETVYKTYTPEAEK 575  
 DB 541 KGVNRKLGRTGLYPSYKVKREKIETVYKTYTPEADK 575

## RESULT 4

US-09-442-629-2  
 ; Sequence 2, Application US/09442629  
 ; Patent No. 6291219  
 ; GENERAL INFORMATION:  
 ; APPLICANT: TANIGUCHI, Naoyuki  
 ; APPLICANT: UOZUMI, Naofumi  
 ; APPLICANT: SHIBA, Tetsuo  
 ; APPLICANT: YANAGIDANI, Shusaku  
 ; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
 ; NUMBER OF SEQUENCES: 15  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Kenyon & Kenyon  
 ; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
 ; CITY: Washington  
 ; STATE: DC  
 ; COUNTRY: US  
 ; ZIP: 20036  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3+ Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
 ; SOFTWARE: WordPerfect 6.1 Windows  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/442.629  
 ; FILING DATE: 18-NO. 6291219-1999  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/913.805A  
 ; FILING DATE: 7 JAN 1998  
 ; APPLICATION NUMBER: PCT/JP97/00171  
 ; FILING DATE: 23 JAN 1997  
 ; APPLICATION NUMBER: JP 192260  
 ; FILING DATE: 22 JUL 1996  
 ; APPLICATION NUMBER: JP 162813  
 ; FILING DATE: 24 JUN 1996  
 ; APPLICATION NUMBER: JP 161648  
 ; FILING DATE: 24 JUN 1996  
 ; APPLICATION NUMBER: JP 10365  
 ; FILING DATE: 24 JAN 1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Toffenetti, Judith L.  
 ; REGISTRATION NUMBER: 39,048  
 ; REFERENCE/DOCKET NUMBER: 2356/3  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 202-429-1776  
 ; TELEFAX: 202-429-0796

```
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 575 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-442-629-2

Query Match          93.6%; Score 2885; DB 3; Length 575;
Best Local Similarity 92.9%; Pred. No. 1,1e-260;
Matches 534; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

QY 1 MEAWTGSNRWIMLILFANGTLLFYGGHLVRDNDHPDHSRELSKILAKLERLKQONEDL 60
DB 1 MRPWTSNRWIMLILFANGTLLFYGGHLVRDNDHSRELSKILAKLERLKQONEDL 60
QY 61 RMAESLRIPESPIQGTATGRVRLVEQLVKAKEQIENYKKQANDLKGDEHILRRRIE 120
DB 61 RMAESLRIPESPIQGPASGRVRALEQFMKAKEQIENYKKQTKNGFGPKGDEHILRRRIE 120
QY 121 NGAKELWFFLQSELKKLKEGNELOHRADETLDDLGHHERSIMTDLVYLSOTDAGGWR 180
DB 121 NGAKELWFFLQSELKKLKEGNELOHRADEFLSDLGHHERSIMTDLVYLSOTDAGGWR 180
QY 181 EKEAKDLTELVRRTIYLNQPKDCSKARKLVNKNKGCGYGCQLHHVYVCFMIAVGTQRT 240
DB 181 EKEAKDLTELVRRTIYLNQPKDCSKAKKLVNKNKGCGYGCQLHHVYVCFMIAVGTQRT 240
QY 241 LLESONWRYATGCHETVPRPVSCTDRSGLSSTGHWSGEVXKQVQVVELPIVDSLHPR 300
DB 241 LALESHNNRYATGCHETVPRPVSCTDRSGSSTGHWSGEVXKQVQVVELPIVDSVHPR 300
QY 301 PYPVLPAVEDLADRLLRVHGDPVAVWVSQFVKYLIRPOPLEREIEETTKLGFKHPVI 360
DB 301 PYPVLPAVEDLADRLLRVHGDPVAVWVSQFVKYLIRPOPLEREIEEATKLGKHPVI 360
QY 361 GVHVRTDKVGTAAFPHPLEIEYMHVVEHFQLLERRMKYDKKRVYLAITDDPSLLKEATK 420
DB 361 GVHVRTDKVGAFAHPHPLEIEYTVHVEEDFQLLARRMQVDKRVYLAITDDPALLKEATK 420
QY 421 YSNYEFISDNSISWSAGLHNRYTENSRLRGVILDIHFLSQADFLVCTFSQVCRVAYEIMQ 480
DB 421 YPSYEFISDNSISWSAGLHNRYTENSRLRGVILDIHFLSQADFLVCTFSQVCRVAYEIMQ 480
QY 481 TLHPDASANFHSIDDIYYPGGNAHQIAVYHPQRTKEEIPMEPGDIIIGVAGNHWGYS 540
DB 481 ALHPDASANFRSDDDIYYPGGNAHQIAVYHPQRTKEEIPMEPGDIIIGVAGNHWGYP 540
QY 541 KGVNRKLGKTLGYPYKYVREKIEYKPYTPAEK 575
DB 541 KGVNRKLGRTGLYPSKYVREKIEYKPYTPAEK 575

RESULT 5
US-08-913-805A-12
; Sequence 12, Application US/08913805A
; Patent No. 6054304
; GENERAL INFORMATION:
; APPLICANT: TANIGUCHI, Naoyuki
; APPLICANT: UOZUMI, Naofumi
; APPLICANT: SHIBA, Tetsuo
; APPLICANT: YANAGIDANI, Shusaku
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: US
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3+ Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
; SOFTWARE: WordPerfect 6.1 Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,805A
; FILING DATE: 7 JAN 1998
; PRIORITY INFORMATION:
; PRIORITY NUMBER: US/08/913,805A
; FILING DATE: 7 JAN 1998
; APPLICATION NUMBER: PCT/JP97/00171
```

```
; MEDIUM TYPE: 3+ Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
; SOFTWARE: WordPerfect 6.1 Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,805A
; FILING DATE: 7 JAN 1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP97/00171
; FILING DATE: 23 JAN 1997
; APPLICATION NUMBER: JP 192260
; FILING DATE: 22 JUL 1996
; APPLICATION NUMBER: JP 162813
; FILING DATE: 24 JUN 1996
; APPLICATION NUMBER: JP 161648
; FILING DATE: 21 JUN 1996
; APPLICATION NUMBER: JP 10365
; FILING DATE: 24 JAN 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Toffenetti, Judith L.
; REGISTRATION NUMBER: 39,048
; REFERENCE/DOCKET NUMBER: 2356/3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-429-1776
; TELEFAX: 202-429-0796
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-913-805A-12

Query Match          4.3%; Score 131; DB 3; Length 25;
Best Local Similarity 100.0%; Pred. No. 1,3e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 KLGFKHPVIGVHVVRTDKVGTAAAF 376
DB 1 KLGFKHPVIGVHVVRTDKVGTAAAF 25

RESULT 6
US-09-442-629-12
; Sequence 12, Application US/09442629
; Patent No. 6291219
; GENERAL INFORMATION:
; APPLICANT: TANIGUCHI, Naoyuki
; APPLICANT: UOZUMI, Naofumi
; APPLICANT: SHIBA, Tetsuo
; APPLICANT: YANAGIDANI, Shusaku
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: US
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3+ Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
; SOFTWARE: WordPerfect 6.1 Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/442,629
; FILING DATE: 18-No. 6291219-1999
; PRIORITY INFORMATION:
; PRIORITY NUMBER: US/08/913,805A
; FILING DATE: 7 JAN 1998
; APPLICATION NUMBER: PCT/JP97/00171
```

;; FILING DATE: 23 JAN 1997  
;; APPLICATION NUMBER: JP 192260  
;; FILING DATE: 22 JUL 1996  
;; APPLICATION NUMBER: JP 162813  
;; FILING DATE: 24 JUN 1996  
;; APPLICATION NUMBER: JP 161648  
;; FILING DATE: 21 JUN 1996  
;; APPLICATION NUMBER: JP 10365  
;; FILING DATE: 24 JAN 1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Toffenetti, Judith L.  
;; REGISTRATION NUMBER: 39,048  
;; REFERENCE/DOCKET NUMBER: 2356/3  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202-429-1776  
;; TELEFAX: 202-429-0796  
;; INFORMATION FOR SEQ ID NO: 12:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 25 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 12:  
US-09-442-629-12  
  
Query Match 4.3%; Score 131; DB 3; Length 25;  
Best Local Similarity 100.0%; Pred. No. 1.3e-05; Length 25;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 352 KLGFKHPVIGVHVRTDKVGTAAAF 376  
Db 1 KLGFKHPVIGVHVRTDKVGTAAAF 25  
  
RESULT 7  
US-09-104-324B-4  
; Sequence 4, Application US/09104324B  
; Patent No. 6232460  
; GENERAL INFORMATION:  
; APPLICANT: T reed, Ozlem; Sahin, Ugur; Pfreundschuh, Michael  
; TITLE OF INVENTION: Methods For Diagnosis And Treating Cancers,  
; TITLE OF INVENTION: And Methods For Identifying Pathogenic Markers In A Sample Of  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Fulbright & Jaworski LLP  
; CITY: New York City  
; STATE: New York  
; ZIP: 10103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage  
; COMPUTER: IBM  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/104,324B  
; FILING DATE: 25-June-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/892,702  
; FILING DATE: 15-July-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hanson, No. 6232460man D.  
; REGISTRATION NUMBER: 30,946  
; REFERENCE/DOCKET NUMBER: LUD 5491  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 318-3000  
; TELEFAX: (212) 752-5958  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 976 amino acids  
; TYPE: amino acid

;; TOPOLOGY: linear  
US-09-104-324B-4  
  
Query Match 3.8%; Score 118.5; DB 3; Length 976;  
Best Local Similarity 20.5%; Pred. No. 0.054; Length 976;  
Matches 90; Conservative 74; Mismatches 164; Indels 111; Gaps 19;  
  
QY 42 ELSKILAKLERLKQKNEDLRMAESLRIPEDQGTATGR---VRVLEEQLVKAKEQIE 98  
Db 435 ELKKVIGKETLLYVKNQFEXIAEELKTEQEL-IGLLQAREKEVHDLLEIQLTAITTSQ 493  
  
QY 99 NYKQARNDLGKHILRRRIENGAKELWFLQSELKKLKGNELOHRADEILLDLGH 158  
Db 494 YYSKEVKD-----LKTELEN-EKLNKLTSHCNKL-SLENKELTQETSDMTLEKN 543  
  
QY 159 HERSI-----MTDLYVLSQTD---GAGSWREKEAKDLTELVRRTTYLQNPKDC 204  
Db 544 QEDINNNKQEEERMLQIENLQETETQLRNELEYVREELQKRDVECKLD--KSEENC 601  
  
QY 205 SKARKLVCNIN-----KCGYGCQLHHVYVCFMAYGTQRTLLLESQNRW 249  
Db 602 NNLRQVENKNKYIHELQENKALKKGTAKQLN--VYEIKV---NKLELELSAQK 656  
  
QY 250 YATGGWETVFRPVSHTCRSLSTGSHWSGEVKQNVQVVELPIVDSLHPPPYLP 309  
Db 657 F-----GEITDTYQKEIEDKKIS-----E 675  
  
QY 310 EDLADRL--LRVHGUPAV-----WVVSQFVKYLIRPQWLEREIEBTTKKLG-F 355  
Db 676 ENLLEVEKAKVIADEAVKLOKEIDKRCQKIAEVALMEKHQYDKIIBERDSEGLY 735  
  
QY 356 KHPVIGVHVRTDKVGTAAAFPIEYVMVHVEEHFOLLERMMKVDKRVLYLATDPSLLK 415  
Db 736 KSK---BOEQSSLRASLELSNLKAEALLSVKQLE-IEREKEKLEKE--AKENTATLK 789  
  
QY 416 EAKTYSNYEFISDHSISW 434  
Db 790 EKDKKKTQTFLETPPIYW 808  
  
RESULT 8  
PCT-US96-08950-2  
; Sequence 2, Application PC/TUS9608950  
; GENERAL INFORMATION:  
; APPLICANT: OSTEOSA INC.  
; TITLE OF INVENTION: Osteoclast Growth Regulatory Factor  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/08950  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Paul C. Steinhart  
; REGISTRATION NUMBER: 30,806  
; REFERENCE/DOCKET NUMBER: FP-ND 2121  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 214 amino acids

[illegible]

**TITLE OF INVENTION:** Nucleic Acid Encoding a Transiently







**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 08:40:23 ; Search time 217 Seconds  
(without alignments)  
550.686 Million cell updates/sec

Title: US-09-971-773-23  
Perfect score: 3081  
Sequence: 1 MRAWTGSWRWIMLILFWANGT.....YKVRKTIETVKYTPYPAEK 575

Scoring table: BLOSUM62

Gapop 10.0 ; Gapext 0.5

Searched: 789580 seqs, 207824079 residues

Total number of hits satisfying chosen parameters: 789580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3081	100.0	575	11	US-09-971-773-23
2	3024	98.1	575	11	US-09-971-773-24
3	2973	96.5	575	9	US-09-839-136-10
4	2878	93.4	575	9	US-09-839-136-2
5	2639	85.7	515	15	US-10-106-698-4658
6	431	14.0	82	9	US-09-864-761-34720
7	425	13.8	81	9	US-09-864-761-46107
8	133	4.3	1959	12	US-10-028-248A-36
9	132	4.3	485	12	US-10-104-047-3419
10	128	4.2	678	12	US-10-369-493-5427
11	127.5	4.1	1959	12	US-10-028-248A-106
12	125	4.1	1961	12	US-10-028-248A-103
13	124.5	4.0	1960	12	US-10-236-031B-62
14	124.5	4.0	1960	12	US-10-028-248A-104
15	124	4.0	405	12	US-10-094-749-3191

16	124	4.0	1285	15	US-10-205-823-273	Sequence 273, App
17	124	4.0	1285	15	US-10-177-293-317	Sequence 317, App
18	122	4.0	25	9	US-09-839-136-12	Sequence 12, Appl
19	121	3.9	1055	12	US-10-369-493-12504	Sequence 12504, A
20	118.5	3.8	976	12	US-10-117-937-596	Sequence 596, App
21	118	3.8	1999	12	US-10-028-248A-107	Sequence 107, App
22	116.5	3.8	1961	12	US-10-028-248A-105	Sequence 105, App
23	116	3.8	229	9	US-09-879-957-221	Sequence 221, App
24	116	3.8	868	10	US-09-884-001-19	Sequence 19, Appl
25	116	3.8	1156	12	US-10-369-493-43	Sequence 43, Appl
26	116	3.8	2383	15	US-10-082-830-260	Sequence 260, App
27	115.5	3.7	888	11	US-09-893-519A-73	Sequence 73, Appl
28	115	3.7	896	12	US-10-369-493-18106	Sequence 18106, A
29	115	3.7	1027	12	US-10-080-608A-27	Sequence 27, Appl
30	115	3.7	1027	12	US-10-370-685-116	Sequence 116, App
31	115	3.7	2326	12	US-10-369-493-6374	Sequence 6374, Ap
32	113.5	3.7	262	12	US-10-032-189-79	Sequence 79, Appl
33	112.5	3.7	672	12	US-10-369-493-7029	Sequence 7029, Ap
34	112.5	3.7	672	12	US-10-369-493-7030	Sequence 7030, Ap
35	112.5	3.7	1979	15	US-10-205-823-419	Sequence 419, App
36	112	3.6	586	12	US-10-097-111-286	Sequence 286, App
37	111.5	3.6	644	12	US-10-289-762-63	Sequence 63, Appl
38	111.5	3.6	880	12	US-10-369-493-21643	Sequence 21643, A
39	111.5	3.6	919	12	US-10-369-493-22825	Sequence 22825, A
40	111	3.6	1137	12	US-10-369-493-6931	Sequence 6931, Ap
41	111	3.6	1441	12	US-10-412-897-3	Sequence 3, Appli
42	111	3.6	1549	12	US-10-369-493-5702	Sequence 5702, Ap
43	111	3.6	1597	14	US-10-017-216-6	Sequence 6, Appli
44	111	3.6	1641	14	US-10-017-216-5	Sequence 5, Appli
45	111	3.6	2055	14	US-10-017-216-4	Sequence 4, Appli

## ALIGNMENTS

## RESULT 1

US-09-971-773-23  
; Sequence 23, Application US/09971773  
; Publication No: US20030115614A1  
; GENERAL INFORMATION:  
; APPLICANT: Yutaka KANDA  
; APPLICANT: Mitsuo SATOH  
; APPLICANT: Kazuyasu NAKAMURA  
; APPLICANT: Kazuhisa UCHIDA  
; APPLICANT: Toyohide SHINKAWA  
; APPLICANT: Naoko YAMANE  
; APPLICANT: Motoo YAMASAKI  
; APPLICANT: NO. US20030115614A1us HANAI  
; TITLE OF INVENTION: ANTIBODY COMPOSITION-PRODUCING CELL  
; FILE REFERENCE: 249-202  
; CURRENT APPLICATION NUMBER: US/09/971,773  
; CURRENT FILING DATE: 2002-08-30  
; PRIOR APPLICATION NUMBER: JP 2000-308526  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 60/268,926  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 23  
; LENGTH: 575  
; TYPE: PRT  
; ORGANISM: Cricetulus griseus  
US-09-971-773-23

Query Match 100.0%; Score 3081; DB 11; Length 575;  
Best Local Similarity 100.0%; Pred. No. 1.4e-266;  
Matches 575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MRAWTGSWRWIMLILFWANGTLLFYIGGHLVRDNDHPDHSSELSKILAKLERLKOQNEDL 60  
Db 1 MRAWTGSWRWIMLILFWANGTLLFYIGGHLVRDNDHPDHSSELSKILAKLERLKOQNEDL 60  
Qy 61 RMAESLRIPGPIDQGTATGRVRLERQLEQVKAKEQIENYKQARNDLGDKDHEILRRRIE 120

Db 61 RMAESLRIPGPIDOGTATGRVRVLEBQVKAKEQIENYKQARNDLGKDHEILRRRIE 120  
Qy 121 NGAKELWFFLOSELKLLKLEGNELQRADETLILDLGHHERSIMTDLYYLSQTDGAGBWR 180  
Db 121 NGAKELWFFLOSELKLLKLEGNELQRADETLILDLGHHERSIMTDLYYLSQTDGAGBWR 180  
Qy 181 EKEAKDLTELVRRIITYLQNPDKCSKARKLVNINCKGCGYCOLHHVVCYCFMIAIYGTORT 240  
Db 181 EKEAKDLTELVRRIITYLQNPDKCSKARKLVNINCKGCGYCOLHHVVCYCFMIAIYGTORT 240  
Qy 241 LILESQNRWYATGHWETVFRPVSETCTDRSGLSGTGHSGEVKNQVQVVELPIVDSLHPR 300  
Db 241 LILESQNRWYATGHWETVFRPVSETCTDRSGLSGTGHSGEVKNQVQVVELPIVDSLHPR 300  
Qy 301 PPYLPLAVPEDLADRLRLRVHGDPAVWVSQFVKYLIRPQWLERIEBETTKLGFKHPVI 360  
Db 301 PPYLPLAVPEDLADRLRLRVHGDPAVWVSQFVKYLIRPQWLERIEBETTKLGFKHPVI 360  
Qy 361 GVHVVRTDKVGTAAAFHPIIEEYVMVHVEHFQLLERMKVDKRVYLATDDPSLLKEAKTK 420  
Db 361 GVHVVRTDKVGTAAAFHPIIEEYVMVHVEHFQLLERMKVDKRVYLATDDPSLLKEAKTK 420  
Qy 421 YSNYEFISDINSISWAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
Db 421 YSNYEFISDINSISWAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
Qy 481 TLHPDASANFHSLLDDIYFVGQNAHNOIAVYPHQPRTKEEIPMEPGDIIGVAGNHNGYS 540  
Db 481 TLHPDASANFHSLLDDIYFVGQNAHNOIAVYPHQPRTKEEIPMEPGDIIGVAGNHNGYS 540  
Qy 541 KGVNRKLGKTGLYPSYKVKREKIEYKTYPTYPEAEK 575  
Db 541 KGVNRKLGKTGLYPSYKVKREKIEYKTYPTYPEAEK 575

## RESULT 2

US-09-971-773-24  
; Sequence 24, Application US/09971773  
; Publication No. US20030115614A1  
; GENERAL INFORMATION:  
; APPLICANT: Yutaka KANDA  
; APPLICANT: Mitsuo SATOH  
; APPLICANT: Kazuyasu NAKAMURA  
; APPLICANT: Kazuhisa UCHIDA  
; APPLICANT: Toyohide SHINKAWA  
; APPLICANT: Naoko YAMANE  
; APPLICANT: Motoo YAMASAKI  
; APPLICANT: No. US20030115614A1uo HANAI  
; TITLE OF INVENTION: ANTIBODY COMPOSITION-PRODUCING CELL  
; FILE REFERENCE: 249-202  
; CURRENT APPLICATION NUMBER: US/09/971,773  
; CURRENT FILING DATE: 2002-08-30  
; PRIOR APPLICATION NUMBER: JP 2000-308526  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 60/268,926  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 24  
; LENGTH: 575  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-971-773-24

Query Match 98.1%; Score 3024; DB 11; Length 575;  
Best Local Similarity 97.6%; Pred. No. 1.8e-261;  
Matches 561; Conservative 9; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 MRWGTGSRWIMLILFANGTLLFYIGGHLVRDNDHPDHSRSLSKILAKLERLKQONEDL 60  
Db 1 MRWGTGSRWIMLILFANGTLLFYIGGHLVRDNDHPDHSRSLSKILAKLERLKQONEDL 60

Qy 61 RMAESLRIPGPIDOGTATGRVRVLEBQVKAKEQIENYKQARNDLGKDHEILRRRIE 120  
Db 61 RMAESLRIPGPIDOGTATGRVRVLEBQVKAKEQIENYKQARNDLGKDHEILRRRIE 120  
Qy 121 NGAKELWFFLOSELKLLKLEGNELQRADETLILDLGHHERSIMTDLYYLSQTDGAGBWR 180  
Db 121 NGAKELWFFLOSELKLLKLEGNELQRADETLILDLGHHERSIMTDLYYLSQTDGAGBWR 180  
Qy 181 EKEAKDLTELVRRIITYLQNPDKCSKARKLVNINCKGCGYCOLHHVVCYCFMIAIYGTORT 240  
Db 181 EKEAKDLTELVRRIITYLQNPDKCSKARKLVNINCKGCGYCOLHHVVCYCFMIAIYGTORT 240  
Qy 241 LILESQNRWYATGHWETVFRPVSETCTDRSGLSGTGHSGEVKNQVQVVELPIVDSLHPR 300  
Db 241 LILESQNRWYATGHWETVFRPVSETCTDRSGLSGTGHSGEVKNQVQVVELPIVDSLHPR 300  
Qy 301 PPYLPLAVPEDLADRLRLRVHGDPAVWVSQFVKYLIRPQWLERIEBETTKLGFKHPVI 360  
Db 301 PPYLPLAVPEDLADRLRLRVHGDPAVWVSQFVKYLIRPQWLERIEBETTKLGFKHPVI 360  
Qy 361 GVHVVRTDKVGTAAAFHPIIEEYVMVHVEHFQLLERMKVDKRVYLATDDPSLLKEAKTK 420  
Db 361 GVHVVRTDKVGTAAAFHPIIEEYVMVHVEHFQLLERMKVDKRVYLATDDPSLLKEAKTK 420  
Qy 421 YSNYEFISDINSISWAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
Db 421 YSNYEFISDINSISWAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
Qy 481 TLHPDASANFHSLLDDIYFVGQNAHNOIAVYPHQPRTKEEIPMEPGDIIGVAGNHNGYS 540  
Db 481 TLHPDASANFHSLLDDIYFVGQNAHNOIAVYPHQPRTKEEIPMEPGDIIGVAGNHNGYS 540  
Qy 541 KGVNRKLGKTGLYPSYKVKREKIEYKTYPTYPEAEK 575  
Db 541 KGVNRKLGKTGLYPSYKVKREKIEYKTYPTYPEAEK 575

## RESULT 3

US-09-839-136-10  
; Sequence 10, Application US/09839136  
; Patent No. US20020081694A1  
; GENERAL INFORMATION:  
; APPLICANT: Naoyuki TANIGUCHI et al.  
; TITLE OF INVENTION: ALPHA 1-6 FUCOSYLTRANSFERASE  
; FILE REFERENCE: 2356-7  
; CURRENT APPLICATION NUMBER: US/09/839,136  
; CURRENT FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: 09/442,629  
; PRIOR FILING DATE: 1999-11-18  
; PRIOR APPLICATION NUMBER: 08/913,805  
; PRIOR FILING DATE: 1998-01-07  
; PRIOR APPLICATION NUMBER: PCT/JP97/00171  
; PRIOR FILING DATE: 1997-01-23  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0.  
; SEQ ID NO 10  
; LENGTH: 575  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-839-136-10

Query Match 96.5%; Score 2973; DB 9; Length 575;  
Best Local Similarity 95.8%; Pred. No. 6.4e-257;  
Matches 551; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

Qy 1 MRWGTGSRWIMLILFANGTLLFYIGGHLVRDNDHPDHSRSLSKILAKLERLKQONEDL 60  
Db 1 MRWGTGSRWIMLILFANGTLLFYIGGHLVRDNDHPDHSRSLSKILAKLERLKQONEDL 60  
Qy 61 RMAESLRIPGPIDOGTATGRVRVLEBQVKAKEQIENYKQARNDLGKDHEILRRRIE 120  
Db 61 RMAESLRIPGPIDOGTATGRVRVLEBQVKAKEQIENYKQARNDLGKDHEILRRRIE 120

Qy 121 NGAKELWFFLOSELKLLKLEGNELQRADEBLLDLGHHSIMTDLYYLSQTDGAGWR 180  
Db 121 NGAKELWFFLOSELKLLKLEGNELQRADEBLLDLGHHSIMTDLYYLSQTDGAGWR 180  
Qy 181 EKEAKDLTELQVRRITTYLQNPDKCSKARKLVNKNKGGCGYCOLHHVYCFMAYGTORT 240  
Db 181 EKEAKDLTELQVRRITTYLQNPDKCSKARKLVNKNKGGCGYCOLHHVYCFMAYGTORT 240  
Qy 241 LILESQNRWYATGWTETFRPVSCTCTDRSGSLSTGHWSGEVKDKNVQVVELPIVDSLHPR 300  
Db 241 LILESQNRWYATGWTETFRPVSCTCTDRSGSLSTGHWSGEVKDKNVQVVELPIVDSLHPR 300  
Qy 301 PPYLPLAVPEDLADRLRLRVHGDPAVMWVSQVVKYLIRPQPMLEREIEETTKLGFKHPVI 360  
Db 301 PPYLPLAVPEDLADRLRVHGDPAVMWVSQVVKYLIRPQPMLEKEIEBATKLGFKHPVI 360  
Qy 361 GVHVRTDKVGEAAFPPIEYVMVHEHFOLERRMKVDKRVYLATDDPSLLKEATK 420  
Db 361 GVHVRTDKVGEAAFPPIEYVMVHEHFOLERRMKVDKRVYLATDDPSLLKEATK 420  
Qy 421 YSNYEFISDINSISWSAGLHNRYSNLSRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
Db 421 YPNYEFISDINSISWSAGLHNRYSNLSRGVILDIHFLSQADFLVCTFSSQVCRVYIEMQ 480  
Qy 481 TLHPDASANFSLDDIYFVGQNAHQIAVYPHQPRTKEEIPMBEPGDIIGVAGNHNWGS 540  
Db 481 TLHPDASANFSLDDIYFVGQNAHQIAVYAHQPRTADEIPMBEPGDIIGVAGNHNWGS 540  
Qy 541 KGVNRKLGKTGLYPSYKVKREKIETKYTYPEAEK 575  
Db 541 KGVNRKLGRTGLYPSYKVKREKIETKYTYPEAEK 575

## RESULT 4

US-09-839-136-2  
; Sequence 2, Application US/09839136  
; Patent No. US20020081694A1  
; GENERAL INFORMATION:  
; APPLICANT: Naoyuki TANIGUCHI et al.  
; TITLE OF INVENTION: ALPHA 1-6 FUCOSYLTRANSFERASE  
; FILE REFERENCE: 2356-7  
; CURRENT APPLICATION NUMBER: US/09/839,136  
; CURRENT FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: 09/442,629  
; PRIOR FILING DATE: 1999-11-18  
; PRIOR APPLICATION NUMBER: 08/913,805  
; PRIOR FILING DATE: 1998-01-07  
; PRIOR APPLICATION NUMBER: PCT/JP97/00171  
; PRIOR FILING DATE: 1997-01-23  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 575  
; TYPE: PRT  
; ORGANISM: Pig  
US-09-839-136-2

Query Match 93.4%; Score 2878; DB 9; Length 575;  
Best Local Similarity 92.7%; Pred. No. 2e-248;  
Matches 533; Conservative 16; Mismatches 26; Indels 0; Gaps 0;  
Qy 1 MRWTGSRWMTMLIFAWGTLLFYIGGHLVRDNDHPDSSRELKILAKLERLKOONEDL 60  
Db 1 MRWTGSRWMTMLIFAWGTLLFYIGGHLVRDNDSDSSRELKILAKLERLKOONEDL 60  
Qy 61 RMMAESLRIPEGPIDQGTATGRVRVLEBQLVKAKEQIENYKQARNDLGKHDEILRRRIE 120  
Db 61 RMAGSLRIPEGPIDQGPASGRVRALEEQFMKAKEQIENYKQTKNGPGKHDEILRRRIE 120  
Qy 121 NGAKELWFFLOSELKLLKLEGNELQRADEBLLDLGHHSIMTDLYYLSQTDGAGWR 180  
Db 121 NGAKELWFFLOSELKLLKLEGNELQRADEBLLDLGHHSIMTDLYYLSQTDGAGWR 180

Qy 181 EKEAKDLTELQVRRITTYLQNPDKCSKARKLVNKNKGGCGYCOLHHVYCFMAYGTORT 240  
Db 181 EKEAKDLTELQVRRITTYLQNPDKCSKARKLVNKNKGGCGYCOLHHVYCFMAYGTORT 240  
Qy 241 LILESQNRWYATGWTETFRPVSCTCTDRSGSLSTGHWSGEVKDKNVQVVELPIVDSLHPR 300  
Db 241 LALESHNRWYATGWTETFRPVSCTCTDRSGSLSTGHWSGEVKDKNVQVVELPIVDSLHPR 300  
Qy 301 PPYLPLAVPEDLADRLRLRVHGDPAVMWVSQVVKYLIRPQPMLEREIEETTKLGFKHPVI 360  
Db 301 PPYLPLAVPEDLADRLRVHGDPAVMWVSQVVKYLIRPQPMLEKEIEBATKLGFKHPVI 360  
Qy 361 GVHVRTDKVGEAAFPPIEYVMVHEHFOLERRMKVDKRVYLATDDPSLLKEATK 420  
Db 361 GVHVRTDKVGEAAFPPIEYTVHVEEDFOLLARRMQVDKRVYLATDDPALLKEATK 420  
Qy 421 YSNYEFISDINSISWSAGLHNRYSNLSRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
Db 421 YPNYEFISDINSISWSAGLHNRYSNLSRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
Qy 481 TLHPDASANFSLDDIYFVGQNAHQIAVYPHQPRTKEEIPMBEPGDIIGVAGNHNWGS 540  
Db 481 ALHPDASANFSLDDIYFVGPNANQIAIYPHQPRTEGEIPMBEPGDIIGVAGNHNWGS 540  
Qy 541 KGVNRKLGKTGLYPSYKVKREKIETKYTYTYPEAEK 575  
Db 541 KGVNRKLGRTGLYPSYKVKREKIETKYTYTYPEADK 575

## RESULT 5

US-10-106-698-4658  
; Sequence 4658, Application US/10106698  
; Publication No. US20030105690A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide  
; FILE REFERENCE: PA005P1  
; CURRENT APPLICATION NUMBER: US/10/106,698  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: PCT/US00/26524  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: US 60/157,137  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: US 60/163,280  
; PRIOR FILING DATE: 1999-11-03  
; NUMBER OF SEQ ID NOS: 8564  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4658  
; LENGTH: 515  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-106-698-4658

Query Match 85.7%; Score 2639; DB 15; Length 515;  
Best Local Similarity 95.7%; Pred. No. 4.1e-227;  
Matches 489; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

Qy 65 ESLRIPEGPIDQGTATGRVRVLEBQLVKAKEQIENYKQARNDLGKHDEILRRRIENGAK 124  
Db 5 KSLRIPEGPIDQGPALGRVRVLEBQLVKAKEQIENYKQTRNGLGKHDEILRRRIENGAK 64  
Qy 125 ELWFFLOSELKLLKLEGNELQRADEBLLDLGHHSIMTDLYYLSQTDGAGWREKEA 184  
Db 65 ELWFFLOSELKLLKLEGNELQRADEBLLDLGHHSIMTDLYYLSQTDGAGWREKEA 124  
Qy 185 KDLTELQVRRITTYLQNPDKCSKARKLVNKNKGGCGYCOLHHVYCFMAYGTORTLILE 244  
Db 125 KDLTELQVRRITTYLQNPDKCSKARKLVNKNKGGCGYCOLHHVYCFMAYGTORTLILE 184  
Qy 245 SQNRWYATGWTETFRPVSCTCTDRSGSLSTGHWSGEVKDKNVQVVELPIVDSLHPRPPYL 304  
Db 185 SQNRWYATGWTETFRPVSCTCTDRSGSLSTGHWSGEVKDKNVQVVELPIVDSLHPRPPYL 244

```
QY 305 PLAVPEDLADLLRVHGDPAVWVWSQFVKYLIRPOFWLERIEETTKLGLGFKHPVIGVH 364
Db 245 PLAVPEDLADLLRVHGDPAVWVWSQFVKYLIRPOFWLERIEETTKLGLGFKHPVIGVH 304
QY 365 RTDKVGTAAAPHTIEEYVHVVEEHFOLLEREMKVDKRVYLATDDPSLLKEAKTKYSNY 424
Db 305 RTDKVGTAAAPHTIEEYVHVVEEHFOLLARMQDKRVYLATDDPSLLKEAKTKYPNY 364
QY 425 EPISNSISWSAGLHNRVTENSLRGVILDDIHFLSQADFLVCTFFSQVCRVAVEIMQTLHP 484
Db 365 EPISNSISWSAGLHNRVTENSLRGVILDDIHFLSQADFLVCTFFSQVCRVAVEIMQTLHP 424
QY 485 DASANFHSLLDDIYFGGNAHNAQIAVYHPQPTKEEIMPEPGDIIIGVAGNHNWGSKGVN 544
Db 425 DASANFHSLLDDIYFGGNAHNAQIAVYHPQPTADEIPMEPGDIIIGVAGNHNWGSKGVN 484
QY 545 RKLGTGLPSYKVKREKIEYKPYTPEAEK 575
Db 485 RKLGTGLPSYKVKREKIEYKPYTPEAEK 515
```

## RESULT 6

US-09-864-761-34720

; Sequence 34720, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

; FILE REFERENCE: Aecomica-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: US 09/774,203

; PRIOR FILING DATE: 2001-01-29

; NUMBER OF SEQ ID NOS: 49117

; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

```
; SEQ ID NO 34720
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL109847.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.9
; OTHER INFORMATION: SWISSPROT HIT: Q9VLZ7, EVALU6 6.80e+00
; OTHER INFORMATION: EST_HUMAN HIT: AW002060.1, EVALU6 7.00e-44
US-09-864-761-34720
```

Query Match 14.0%; Score 431; DB 9; Length 82;

Best Local Similarity 96.3%; Pred. No. 8e-31;

Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 280 EVKDKNVQVVELPIVDSLHPRPPYPLAVPEDLADLLRVHGDPAVWVWSQFVKYLIRPQ 339

Db 1 EVKDKNVQVVELPIVDSLHPRPPYPLAVPEDLADLLRVHGDPAVWVWSQFVKYLIRPQ 60

QY 340 PWLERIEETTKLGLGFKHPVIG 361

Db 61 PWLSKEIEEATKLGFKHPVIG 82

## RESULT 7

US-09-864-761-46107

; Sequence 46107, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

; FILE REFERENCE: Aecomica-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30



```
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 46107
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL109847.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
; OTHER INFORMATION: SWISSPROT HIT: Q9VLZ7, EVALUATION 6.40e+00
; OTHER INFORMATION: EST_HUMAN HIT: AW002060.1, EVALUATION 5.00e-43
US-09-864-761-46107

Query Match 13.8%; Score 425; DB 9; Length 81;
Best Local Similarity 96.3%; Pred. No. 2.7e-30;
Matches 78; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 280 EVKKNVQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVVSQFVKYLIRPQ 339
DB 1 EVKKNVQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVVSQFVKYLIRPQ 60

QY 340 PWLEIEIETTKLGFKEHPVI 360
DB 61 PWLEKEIEEATKUGFKHPVI 81

RESULT 8
US-10-028-248A-36
; Sequence 36, Application US/10028248A
; Publication No. US20030235882A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Patturajan, Meera
; APPLICANT: Vernet, Corine
; APPLICANT: Casman, Stacie
; APPLICANT: Malyankar, Uriel
; APPLICANT: Shenoy, Suresh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gangolli, Esha
; APPLICANT: Miller, Charles
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Smithson, Glennda
; APPLICANT: Zerkhusen, Bryan
; APPLICANT: Liu, Xiaohong
; APPLICANT: Colman, Steven
; APPLICANT: Tchernev, Velizar
; APPLICANT: Si, Jingsheng
; APPLICANT: Edinger, Shlomit
; APPLICANT: Stone, David
; APPLICANT: Sciore, Paul
; APPLICANT: Millet, Isabelle
; APPLICANT: Rothenberg, Mark
; TITLE OF INVENTION: No. US20030235882A1el Nucleic Acids and Polypeptides and Methods
; FILE REFERENCE: 21402-222
; CURRENT APPLICATION NUMBER: US/10/028,248A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256619
; PRIOR FILING DATE: 2000-12-19
```

```
; PRIOR APPLICATION NUMBER: 60/262959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311266
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 211
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 36
; LENGTH: 1959
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-028-248A-36

Query Match 4.3%; Score 133; DB 12; Length 1959;
Best Local Similarity 15.1%; Pred. No. 0.041;
Matches 105; Conservative 95; Mismatches 191; Indels 160; Gaps 25;

QY 32 DNDPHDSSRELKILAKLERLKQO---NEDLARMAESLRIPGPDIDQGTATGVRVLE 87
DB 1054 EGDSTDLS-D-QIAELQAIKELKMLAKKEBELQ--AALARVEEAAQKNWALKKIRELE 1110
QY 88 EQLVAKKEQIB-----NYKQARNDLGKHDEILRRRIENGAKELWFFLQSELKKLKLE 141
DB 1111 SQISELQEDLESEASRNKAEKQKRDGEELEALKTELEDTLDSI--AAQQLSRKQEQ 1168
QY 142 GNEIQRHADEILLDLGHHERSINTDLYLVSOTDGAGWEKREKADLTVELVORRITYLQNP 201
DB 1169 VNILKKTLEE--EAKTHEAQIQ-----EMRQKHSQAVEELAEQ----- 1204
QY 202 KDCSKARKLVNINIKGCGYGCQLHHVYVCFMIAGTQRTLLILESONWRYATGGWETVRFP 261
DB 1205 --LEQTKRVKANLEK-----AKQTL-----ENER----- 1226
QY 262 VSETCTDRSLTCHWSGEVKNV--QVVELPIVDSLHPRPPYPLAVPEDLADRLRV 319
DB 1227 -GELANVKKVLLQGRSEHKKVKAQQLQVKFNEGR-----VTELADKVTK- 1277
QY 320 HGDPVWVVSQFVKYLIRPQPLEREIEBETTKLGFKEHPVIGVHVRRTDKVGTAA--FH 377
DB 1278 -----LQVELDNVTGLLS-----QSDSKSKLTQDFS 1304
QY 378 PIEYMWVVEHFQLLERRMKVKKRYLATDDPSLLKEAKTKYSNYEFFISNLSWSAG 437
DB 1305 ALESQLODTQELLQ-----EENRQKLSLTK-----LQO-----VEDEKNSFRBQ 1344
QY 438 LHNRYTENSRLRGVILDIHFLSQADFLVCTFSQVC-RVAYEIMOTLHPDASANPHSLDDI 496
DB 1345 LEEBEAKHNLEKQIATLH-AQVADMKKMEVSVCLETAEEVKRKLQD-----LEGL 1396
QY 497 YFYGQNAHQIAVYVPHQPTKEIPIPEPGDIIIGVAGNHMNGYSKGNVRKLGKTGLPSY 556
DB 1397 ----SQRHEEKVAAVDKLEKTKRLQOELDLL--VDLDHQRSACNLEKKQK---FDQL 1448
QY 557 KVREKIETVKY 567
DB 1449 LAEKEKTSISKY 1459

RESULT 9
US-10-104-047-3419
; Sequence 3419, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2003-03-25
```

; PRIOR APPLICATION NUMBER:

; PRIOR FILING DATE:

; NUMBER OF SEQ ID NOS: 4096

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3419

; LENGTH: 485

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-104-047-3419

Query Match 4.3%; Score 132; DB 12; Length 485;

Best Local Similarity 20.5%; Pred. No. 0.0063;

Matches 93; Conservative 72; Mismatches 137; Indels 152; Gaps 23;

QY 45 KILAKLERLK-QONEDLRMAESLRIPGPIDQGTATGRVRLVLEBQLVKAKEQIENYKQ 103

DB 13 KVTYKAERLKLQEEERLKEE-----EASLK-----YEKE 45

QY 104 ARNDLGKDHILRRRIENGAKELWFFLOSELKKLKKLEGNELQRADEI---LLDLG-- 157

DB 46 -----EMERLEIQRTEKEW-----HRLKADLERNEELEEYLLERCFP 86

QY 158 -----HHERSIMTDLYLSOTDAGE-----WREKADLTTELVR----- 193

DB 87 EAEKLUKQETKLSQWKHYIQCDGSPDPSPVAQEMNTFISLWKEKTNETFEVIEKSKVVLN 146

QY 194 -----RITVLOPKDCSRKARKLVNINCKGCGYCOLHHVVCFMAYGTRTILLESQNW 248

DB 147 LIEKLUKFLILETP-PCDLQDKNIIQYQESI---LQLELH---LKFGVATEILKQ--- 196

QY 249 RYATGGWETVPRVSETCTDRSLSTGHWSGVKDKNVQVVELPITVDSLHPRPPYPLAV 308

DB 197 -----ASTLAD---LDSGNKEVIKDEN---VTLYVWNLKKNPRHRSVRF 236

QY 309 PE-----DLADRLLRVHGD-----PAVWV-----VSQFVKYLIR-PQP 340

DB 237 SETQIGFEIPRILATSDIAVRLHLTHYDVSALHPVTPSPKEYTSVAVTELKDDVKNVEK 296

QY 341 WLERIEETTKLGFKHPVIGVHVHVRTDKVGEAAFPHPTEEYVMVHVEEHFOLL--ERRMK 398

DB 297 AISKVEEESKO-----QERGSHLQEBEIKVEEEOGDIENVMSAARESEAIKICEREMK 351

QY 399 VDKKRVLYATDDPSLLKKAETKYNYSFISDMSI 432

DB 352 VLSETVSA--QLLLVENSSE--KPDPEFEDNVV 380

RESULT 10

US-10-369-493-5427

; Sequence 5427, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 5427

; LENGTH: 678

; TYPE: PRT

; ORGANISM: Caenorhabditis elegans

US-10-369-493-5427

Query Match

Best Local Similarity 4.2%; Score 128; DB 12; Length 678;

Pred. No. 0.024;

Matches 91; Conservative 82; Mismatches 186; Indels 108; Gaps 19;

QY 38 HSSRELSKILAKLERLKQONEDLRMAESLRIPGPIDQGTATGRVRLVLEBQLVKAKEQI 97

DB 255 HUNGMSDIELEASKMEQADEMRSLAEV-----FNERRAKEBELQGLEIEV 302

QY 98 ENYK-----KQARNDLGKDHEILRRR-TENGAKELWFFLOSELKKLK-KLE 141

DB 303 EEOKKLNEAVTHAMPQPMKEYEDLKSEAKLLRERVVEMEAKN--EDLDDRISKYEIEIR 360

QY 142 GNEIQRHA--DEILLDLGHHERSIMTDLYLSOTDAGAGWEKEKADLTTELVRRIYTL 198

DB 361 SNPUKKKAIQIQTDLTKQEEKLMEDMQSALTPEA---WRDKMSNMKQNLNADLVVIE 417

QY 199 QNPK-----DCSKARKLVNINCKGCGYCOLHHVVCFMAYGTRTILLES 245

DB 418 KQHKTVKQDOISLASEELHEVDSQGEAQIMAHHTKYLLDLSSKSTMLDDTTTENVPQIVYQ 477

QY 246 QNRYATGGWETVPRVSE-----TCTDRSGLS--TGHWSGVKDKNVQV--- 288

DB 478 QDIEEFSDAVVLLIRKISANLKKVNLQEDQITDLDERGLTLQTN-VDELKEMHVRLOBEL 536

QY 289 -----VELPI---VDSLHPRPPYPLAVPEDLADRLLRVHGDPAVWVVSQFVKYLIRPQP 340

DB 537 ISIDMELALNEEIDNLETEK-----KIDQELAGVGNVDS----- 574

QY 341 WLERIEETTKLGFKHPVIGVHVHVRTDK--VGTEAAFPHPTEEYVMVHVEEHFOLLERMK 398

DB 575 GLRQLEERQKRLDEAPARSHQMLEANVASIRNELHSIPGYSQH-----KMLRERLE 629

QY 399 VDKKRVLYATDDPSLLKKAETKYNYSFISDMSISWSAGLHNRVTEN 445

DB 630 AVEKRTAAKSUDMSLRKTE-----IDYEDIKTESIRLOEYNNMLLTN 672

RESULT 11

US-10-028-248A-106

; Sequence 106, Application US/10028248A

; Publication No. US200302335882A1

; GENERAL INFORMATION:

; APPLICANT: Shinketsu, Richard

; APPLICANT: Patturajan, Meera

; APPLICANT: Vernet, Corine

; APPLICANT: Casman, Stacie

; APPLICANT: Malyankar, Uriel

; APPLICANT: Shenoy, Suresh

; APPLICANT: Spytek, Kimberly

; APPLICANT: Gangoli, Esha

; APPLICANT: Miller, Charles

; APPLICANT: Boldog, Ferenc

; APPLICANT: Li, Li

; APPLICANT: Taupier Jr, Raymond J

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Smithson, Glennda

; APPLICANT: Zernhusen, Bryan

; APPLICANT: Liu, Xiaohong

; APPLICANT: Colman, Steven

; APPLICANT: Tchernev, Velizar

; APPLICANT: Si, Jingsheng

; APPLICANT: Edinger, Shlomit

; APPLICANT: Stone, David

; APPLICANT: Sciore, Paul

; APPLICANT: Millet, Isabelle

; APPLICANT: Rothenberg, Mark

; TITLE OF INVENTION: No. US200302335882A1el Nucleic Acids and Polypeptides and Methods

; FILE OF INVENTION: Thereof

; FILE REFERENCE: 21402-222

; CURRENT APPLICATION NUMBER: US/10/028,248A

; CURRENT FILING DATE: 2001-12-19

; PRIOR APPLICATION NUMBER: 60/256619

; PRIOR FILING DATE: 2000-12-19

; PRIOR APPLICATION NUMBER: 60/262959

; PRIOR FILING DATE: 2001-01-19

;; PRIOR APPLICATION NUMBER: 60/272408  
;; PRIOR FILING DATE: 2001-02-28  
;; PRIOR APPLICATION NUMBER: 60/285189  
;; PRIOR FILING DATE: 2001-04-20  
;; PRIOR APPLICATION NUMBER: 60/308039  
;; PRIOR FILING DATE: 2001-07-26  
;; PRIOR APPLICATION NUMBER: 60/311266  
;; PRIOR FILING DATE: 2001-08-09  
;; NUMBER OF SEQ ID NOS: 211  
;; SOFTWARE: Patent in Ver. 2.1  
;; SEQ ID NO 106  
;; LENGTH: 1959  
;; TYPE: PRT  
;; ORGANISM: Gallus gallus  
US-10-028-248A-106

Query Match 4.1%; Score 127.5; DB 12; Length 1959;  
Best Local Similarity 19.8%; Pred. No. 0.13;  
Matches 110; Conservative 79; Mismatches 177; Indels 189; Gaps 22;  
QY 42 ELSKILAKLE-RLKQONEDLRMAESLRIPEGPIDQGTATGRVVRVLEQLVKAKEQIE-- 98  
DB 1066 ELQAIAELKQLSKSEBELQ--AALARVEEAAQKNWALKKIRELESQITELQEDLESE 1123  
QY 99 ----NYKQARNDLGKHDEILRRRIEN-----GAKELWFFLQSELKKLKLEGNELQKH 148  
DB 1124 RASRNKAQKQKDLGEELEALKTELEDITLDSAAQQLRSKREQEVTVLVKXLEDAKTH 1183  
QY 149 ADEILLDLGHHERSIMTDLVYLSQTDGAGWEKEAKDLTELVRRIYQLNPQDCSKAR 208  
DB 1184 EAQI-----QEMRQKHSQAIEELAEQ-----LEQTK 1209.  
QY 209 KLVCNKNGCGYGCQLHHVYCFMAYGTQRTLLLESQNWRYATGWTVPVSECTD 268  
DB 1210 RVKANLEKA-----KQALESER-----AELSN 1232  
QY 269 RSGSLTGHWSGVRKDNV--QVVELPIVDSLHPRPPYLPVAPEDLADRL--LRVHGDP 324  
DB 1233 VKVLLOGKDAHKKKVDQALQELQVKTETGER-----VKTELAERVNKLQVELDNV 1285  
QY 325 VWWSQ-----FVKYLIROPWLERI-----EETTKLGFHPVIGVHVRRTDKVGT 372  
DB 1286 TGLLNQSDSKSIK-LAKDFSALQSQDQTELLQEBETRLKLSFS-----TKLKQT 1334  
QY 373 EAAPFPIEBYVHVEEHQLLERRMKVKKRVYLDTPSLKKAETKYSNVEFTSDNSI 432  
DB 1335 EDEKNALKEQEEEEEKRNLEKQISVLQQQAV-----EARKKMD----- 1374  
QY 433 SWSAGLHNRYTENSRLRGVILDIHFLSQADFVLCFTFSSQVCRVAYEIMOTLHPDASANFHS 492  
DB 1375 -----DGLGC-----LEIAEAKKLQKD-----LES 1396  
QY 493 LDDIYFGQNAHQIAYVPHQPTKEEIPMEPGDIIGVAGNHWNYSKGVNRKLGKTGL 552  
DB 1397 LTQRY-----EKKIAYDKLEKTKRLQQLDLD--IAVDLHQRTQVSNLEKKQK--- 1445  
QY 553 YPSYKVRKEIETVY 567  
DB 1446 FDQLAAEKNISAKY 1460

RESULT 12  
US-10-028-248A-103  
; Sequence 103, Application US/10028248A  
; Publication No. US20030235882A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Vernet, Corine  
; APPLICANT: Casman, Stacie  
; APPLICANT: Malyankar, Uriel  
; APPLICANT: Shenoy, Suresh  
; APPLICANT: Spytek, Kimberly

;; APPLICANT: Gangolli, Esha  
;; APPLICANT: Miller, Charles  
;; APPLICANT: Boldog, Ferenc  
;; APPLICANT: Li, Li  
;; APPLICANT: Taupier Jr, Raymond J  
;; APPLICANT: Kekuda, Ramesh  
;; APPLICANT: Smithson, Glennda  
;; APPLICANT: Zertusen, Bryan  
;; APPLICANT: Liu, Xiaohong  
;; APPLICANT: Colman, Stevan  
;; APPLICANT: Tchernev, Velizar  
;; APPLICANT: Si, Jingsheng  
;; APPLICANT: Edinger, Shlomit  
;; APPLICANT: Stone, David  
;; APPLICANT: Sciore, Paul  
;; APPLICANT: Millet, Isabelle  
;; APPLICANT: Rothenberg, Mark  
;; TITLE OF INVENTION: NO. US20030235882A1el Nucleic Acids and Polypeptides and Methods  
;; FILE REFERENCE: 21402-22;  
;; CURRENT APPLICATION NUMBER: US/10/028,248A  
;; PRIOR FILING DATE: 2001-12-19  
;; PRIOR APPLICATION NUMBER: 60/256619  
;; PRIOR FILING DATE: 2000-12-19  
;; PRIOR APPLICATION NUMBER: 60/262959  
;; PRIOR FILING DATE: 2001-01-19  
;; PRIOR APPLICATION NUMBER: 60/272408  
;; PRIOR FILING DATE: 2001-02-28  
;; PRIOR APPLICATION NUMBER: 60/285189  
;; PRIOR FILING DATE: 2001-04-20  
;; PRIOR APPLICATION NUMBER: 60/308039  
;; PRIOR FILING DATE: 2001-07-26  
;; PRIOR APPLICATION NUMBER: 60/311266  
;; PRIOR FILING DATE: 2001-08-09  
;; NUMBER OF SEQ ID NOS: 211  
;; SOFTWARE: Patent in Ver. 2.1  
;; SEQ ID NO 103  
;; LENGTH: 1961  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-028-248A-103

Query Match 4.1%; Score 125; DB 12; Length 1961;  
Best Local Similarity 19.2%; Pred. No. 0.21;  
Matches 106; Conservative 94; Mismatches 191; Indels 162; Gaps 26;  
QY 32 DNDHPDSSRELSILAKLERLKOQ-----NEDLRMAESLRIPEGPIDQGTATGRVVRLE 87  
DB 1054 EGDSTDLSL-QIAELQAIKELKQAKKEBELQ--AALARVEEAAQKNWALKKIRELE 1110  
QY 88 EQLVKAKEQIE-----NYKQARNDLGKHDEILRRRIENGAKELWFFLQSELKKLKLE 141  
DB 1111 SQISELOEDLSEERASRNKAQKQKRDGLGEELEALKTELEDITLDS--AAQQLRSKREOE 1168  
QY 142 GNEQLQHADEILLDLGHHERSIMTDLVYLSQTDGAGWEKEAKDLTELVRRIYQLNP 201  
DB 1169 VNILKXLEE--FAKTHEAQIQ-----EMRQKHSQAIEELAEQ----- 1204  
QY 202 KDCSKARKLVNKNKNGCGYGCQLHHVYCFMAYGTQRTLLLESQNWRYATGWTVPVFP 261  
DB 1205 --LEQTKRVKANLEK-----AKQTL-----ENER----- 1226  
QY 262 VSEICTDRSGLSTGHWSGVRKDNV--QVVELPIVDSLHPRPPYLPVAPEDLADRLRV 319  
DB 1227 -GELANEVYKVLLOGGRDSEHKRKKVEAQQLQELQVKNFNEGER-----VRTEADKVTX- 1277  
QY 320 HGDPVAVWVSQFVKYLIROPWLERIEIETTCKLGFHPVIGVHVRRTDKVGTAA--FH 377  
DB 1278 -----LQVELDNVTGLS-----QSDSKSKLTKDFS 1304  
QY 378 PTEYMHVHEEHFQLLERRMKVKKRVYLDTPSLKKAETKYSNVEFTSDNSISWAG 437  
DB 1305 ALESQLODTQELLQ-----EENRQKLSLSTK-----LQK-----VEDEKNSFREQ 1344



```
Qy 88 EQLVKAQEIE-----NYKQARNDLGKQHEILLRRIENGAKELWFFLQSELKLLKLE 141
Db 1111 SISELQEDLESERAKAEKQKRDGLGEELEALKTELEDTLDTST--AAQQLRSKREQE 1168
Qy 142 GNEIQRHADEILLDLGHHHSIMTDLVYLSOTDAGAGWEKEAKDLTELVRRTYILQNP 201
Db 1169 VNILKKTLEE---EAKTHEAQIQ-----EMRQKHSQAVBELAQ-----1204
Qy 202 KDCSKARKLVNINKGCGYGCQLHHVVYCFMAYGTORTLILESONWRVATGGWETVFRP 261
Db 1205 --LEQTKRVKANLEK-----AKOTL-----ENER-----1226
Qy 262 VSETCTDRSGLSGTGHSGEVQKQNV--QVVELPIVDSLHPRPPYLPPLAVPDLADRLLRV 319
Db 1227 -GELANEVVLQKGDSEHKKVKAQQLQELQVKNFEGE-----VRTELADKVTK- 1277
Qy 320 HGDPVWVWSFVKYLRPQWLEREIEETTKLGFHPVIGVHVRTDKVGTAA--FH 377
Db 1278 -----LOVELDNVTGLLS-----QSDSKSSKLTDKDFS 1304
Qy 378 PIEEYVHVESHFOLLERMKVVKRVYLATDDPSLLKEAKTKYSNYEFISDINSISAG 437
Db 1305 ALESQLOQTOLLO-----BENQKJSLSTK-----VEDEKNSFREQ 1344
Qy 438 L-HNRYTENSIRGVILDIHFLSQADFLVCTFSSQVC-RVAYEIMQTLHPDASANFHSLDD 495
Db 1345 LEEBEEAKHNELEKQIATLH-AQVADMKKKMEDSVGCLTAEVKKLQKD-----LEG 1396
Qy 496 IYFYGQNAHQIAVYHPQRTKBEIPMEPGDITGVAGNHWGYSKGVNRKLGKTGLYPS 555
Db 1397 L-----SORHEEKVAAYDKLETKTRLQELDDLL-VLDLHQRSQACNLEKQKK---PDQ 1448
Qy 556 YKVBKEIETVKY 567
Db 1449 LLAEEKTISAKY 1460
```

## RESULT 15

```
US-10-094-749-3191
; Sequence 3191, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIG, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3191
; LENGTH: 405
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
US-10-094-749-3191
Query Match 4.08; Score 124; DB 12; Length 405;
Best Local Similarity 21.9%; Pred. No. 0.025;
Matches 79; Conservative 54; Mismatches 109; Indels 118; Gaps 18;
Qy 42 ELSKILAKLERLKQONEDLRMAESLRIPGPDIDQGTATGRVVRVLEEQLVKAQEIQENYK 101
Db 116 ELSKY---KEIINRQKAEIQNLDDKVK---TADQ-----LQEQLRGKQEIENLK 159
Qy 102 KQAR-----NDLQKQHEILLRRIENGAKELWFFLQSELKLLKKL--EGNELQRHADEIL 153
Db 160 EEVESLNSLINDLQKDIEGSRKR-----ESELLLFTERLTSKNAQLQSESNSLQSQFDKVS 215
Qy 154 LDLGHHERSIMTDLVYLSOTD-----GAGWEKEAKDL-----187
Db 216 CS-----ESQLOSQCEQMKQTNINLESRLLEEELRKEEVQTLQAEACRQTEVKALSTQV 271
Qy 188 ----TELVRRTYILQNPKDCSK-----ARKLVNINKGCGYGCQLHHVVYCFMAYGTQR 239
Db 272 EELKDELVTQRRKGIASSIKDLTKQLQARRKLDQVESH-----SYDKE- 314
Qy 240 TLILESQNRVATYGTWETVFRPVSETCTDRLSGTGHWSGEVKQKNVQVVELPIVDSLHP 299
Db 315 ---VSSMGSRSSSG---SLNARSSAEDRSPTGT-----SSVAVDNFPQVQDK--- 356
Qy 300 RPPYLPPLAVPEDLADRLLRVHG--DPAVWVWSQFVKYLRPQWLEREIEETTKLGFPH 357
Db 357 -----AMLTIERVRLKKAHARKNEKTEFMEHDHKKQLV-----BEIRKTKKI-YKH 401
Search completed: February 2, 2004, 08:52:03
Job time : 219 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 08:39:43 ; Search time 17 Seconds  
(without alignments)  
3252.764 Million cell updates/sec

Title: US-09-971-773-23  
Perfect score: 3081  
Sequence: 1 MRAWTGSWRWIMLILFAWG.....YKVRKETVKYPTYPEAEK 575

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 76: \*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2984	96.9	575	JC5432	glycoprotein 6- $\alpha$ p
2	995	32.3	818	T32154	hypothetical prote
3	133	4.3	1173	T43527	sp8 protein - f1ss
4	133	4.3	1727	T50073	myosin-like coiled
5	128	4.2	678	H88187	protein C18H9.8 [i
6	127.5	4.1	1959	A33977	myosin heavy chain
7	127	4.1	946	S28061	SCP1 protein - rat
8	125	4.1	1961	A61231	myosin heavy chain
9	124.5	4.0	845	T148176	synaptonemal compl
10	124	4.0	1025	H86250	hypothetical prote
11	124	4.0	2442	T08621	centrosome associa
12	121.5	3.9	3187	JC5837	364K Golgi complex
13	120.5	3.9	978	A70387	conserved hypothet
14	119.5	3.9	1270	T09194	adaptor protein in
15	119.5	3.9	2007	B43402	myosin heavy chain
16	119	3.9	1871	D96796	probable heat choc
17	118.5	3.8	1427	S22695	restin - human
18	118	3.8	581	S46327	gene cel IF A3 pro
19	118	3.8	1999	S21801	myosin heavy chain
20	117.5	3.8	1282	JE0120	glycoprotein A - m
21	116	3.8	993	S49461	synaptonemal compl
22	116	3.8	1156	B70356	chromosome assembl
23	115	3.7	1027	S37711	kinesin heavy chain
24	115	3.7	1976	A59252	myosin heavy chain
25	115	3.7	2326	T29140	hypothetical prote
26	114.5	3.7	503	S23741	hypothetical TPR/T
27	114.5	3.7	1392	A43336	microtubule-vesicl
28	114.5	3.7	2017	A36014	myosin heavy chain
29	114.5	3.7	2057	S61477	myosin II heavy ch

30	113	3.7	435	2	D82905	hypothetical prote
31	112.5	3.7	672	2	T22507	hypothetical prote
32	112.5	3.7	1938	1	JK0178	myosin heavy chain
33	112	3.6	465	2	AE1568	hypothetical prote
34	112	3.6	532	2	C83696	hypothetical prote
35	112	3.6	955	2	S24348	myosin heavy chain
36	111.5	3.6	398	2	T04501	hypothetical prote
37	111.5	3.6	642	2	R86496	hypothetical prote
38	111.5	3.6	642	2	F81543	hypothetical prote
39	111.5	3.6	880	2	F75103	conserved hypothet
40	111.5	3.6	919	2	B72765	hypothetical prote
41	111.5	3.6	1023	2	T33338	hypothetical prote
42	111	3.6	853	2	T51505	hypothetical prote
43	111	3.6	1137	2	T19414	hypothetical prote
44	111	3.6	1549	2	T21809	hypothetical prote
45	111	3.6	1597	2	S68420	citron - mouse

ALIGNMENTS

RESULT 1

JC5432

glycoprotein 6- $\alpha$ -L-fucosyltransferase (EC 2.4.1.68) precursor - human

N;Alternate names: alpha 1-6 FucT; GDP-L-fucose:N-acetyl-beta-D-glucosaminyl:6- $\alpha$ -L-f

C;Species: Homo sapiens (man)

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 16-Jun-2000

A;Accession: JC5432; PC4322

R;Yanagidani, S.; Uozumi, N.; Ihara, Y.; Miyoshi, E.; Yamaguchi, N.; Taniguchi, N.

J. Biochem. 121, 626-632, 1997

A;Title: Purification and cDNA cloning of GDP-L-Fuc:N-acetyl-beta-D-glucosaminide:alpha

A;Reference number: JC5432; MUID:97279058; PMID:9133635

A;Accession: JC5432

A;Molecule type: mRNA

A;Residues: 1-575 <YAN1>

A;Cross-references: DDBJ:D89289; NID:G2055306; PIDN:BAA19764.1; PID:G2055307

A;Accession: PC4322

A;Molecule type: protein

A;Residues: 68-87;352-376;419-432 <YAN2>

C;Comment: This enzyme catalyzes the transfer of fucose from GDP-fucopyranoside to aspar

C;Genetics:

A;Gene: GDB:FUT8

A;Cross-references: GDB:9786294; OMIM:602589

A;Map position: 14q23-14q23

C;Superfamily: human glycoprotein 6- $\alpha$ -L-fucosyltransferase

C;Keywords: glycosyltransferase; hexosyltransferase

F;1-22/Domain: signal sequence #status predicted <SIG>

F;23-575/Product: glycoprotein 6- $\alpha$ -L-fucosyltransferase #status predicted <MAT>

Query Match 96.9%; Score 2984; DB 1; Length 575;  
Best Local Similarity 96.2%; Pred No. 2.2e-186; Mismatches 10; Indels 0; Gaps 0;  
Matches 553; Conservative 10;

Qy	1	MRAWTGSWRWIMLILFAWGTLFFYIGGHLVRDNDHPDHSRSLSKILAKLERLKQONEDL	60
Db	1	MRPTGSWRWIMLILFAWGTLFFYIGGHLVRDNDHPDHSRSLSKILAKLERLKQONEDL	60
Qy	61	RMMAESLRIPGPIIDQGTATGVRVLEBQLVKAQEIENYKQARNDLGKHLEILRRRIE	120
Db	61	RMMAESLRIPGPIIDQGTATGVRVLEBQLVKAQEIENYKQARNDLGKHLEILRRRIE	120
Qy	121	NGAKELPFFLOSELKLLKLGNELQRHADIILLDLGHHSRIMTDLYLSQTDGAGBWR	180
Db	121	NGAKELPFFLOSELKLLKLGNELQRHADIILLDLGHHSRIMTDLYLSQTDGAGBWR	180
Qy	181	EKEAKDTELQVRRITLYLQNPDKCSKARKLYCNINKKCGYCOLHHVYCFMAYGTQRT	240
Db	181	EKEAKDTELQVRRITLYLQNPDKCSKARKLYCNINKKCGYCOLHHVYCFMAYGTQRT	240
Qy	241	LILSQNWRVATGGWETVFRPVSETCTDRSLGISTGHSGEVKDKNVQVVELPVDLSLHPR	300
Db	241	LILSQNWRVATGGWETVFRPVSETCTDRSLGISTGHSGEVKDKNVQVVELPVDLSLHPR	300

QY 301 PPYLPLAVPEDLADRLRVHGDPAVMWSQFVKYLIRPQWLERIEETTKKLGKHPVI 360  
DB 301 PPYLPLAVPEDLADRLRVHGDPAVMWSQFVKYLIRPQWLERIEETTKKLGKHPVI 360  
QY 361 GVHVRTDKVCTEAPHPTEEMVHVEHFQLLERMKVVKRVYLATDDPSLLKEATK 420  
DB 361 GVHVRTDKVCTEAPHPTEEMVHVEHFQLLERMKVVKRVYLATDDPSLLKEATK 420  
QY 421 YSNYEFISDINSISWSAGLHNRNTENSLRGVILDIHFLSOADFLVCTFSQVCRVAYEIMQ 480  
DB 421 YNRYEFISDINSISWSAGLHNRNTENSLRGVILDIHFLSOADFLVCTFSQVCRVAYEIMQ 480  
QY 481 TLHPDASANFHSLLDIYFYGQNAHQAIVYHPQRTKEEIPMEPGDIIGVAGNHNWGS 540  
DB 481 TLHPDASANFHSLLDIYFYGQNAHQAIVYHPQRTKEEIPMEPGDIIGVAGNHNWGS 540  
QY 541 KGVNRKLGRTGLYPSYKVKREKTIETVKYPTYPAEK 575  
DB 541 KGVNRKLGRTGLYPSYKVKREKTIETVKYPTYPAEK 575

RESULT 2  
T32154  
hypothetical protein C10F3.6 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
R:Davidson, S.; Wohlmann, P.  
submitted to the EMBL Data Library, September 1997  
A:Description: The sequence of C. elegans cosmid C10F3.  
A:Reference number: Z21127  
A:Accession: T32154  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-818 <D>  
A:Cross-references: EMBL:AF022968; PIDN:AAB69888.1; GSPDB:GN00023; CESP:C10F3.6  
A:Experimental source: strain Bristol N2; clone C10F3  
C:Genetics:  
A:Map position: 5  
A:Introns: 30/1; 72/2; 235/2; 258/3; 285/2; 315/3; 457/3; 491/1; 544/3; 556/1; 626/1; 64

Query Match 32.3%; Score 995; DB 2; Length 818;  
Best Local Similarity 36.7%; Pred. No. 7.4e-57;  
Matches 213; Conservative 116; Mismatches 188; Indels 64; Gaps 11;  
QY 3 AWTGSRWIMLILPAWGTLLFYIGHLVRDNDPHSHSRELSKILAKLERLKOQNEDLRR 62  
DB 6 AAVGTVMWMTFLYSQ-----SNNQSGDSIRAWRQTKEADIKLOEQNEDLKS 56  
QY 63 MAESLRIPEGPIDQGTATGRVRLBEQLVKAKEQIENYKQARN----- 107  
DB 57 ILEK-----QERNDQHKINEQSHQLPPNPENSLPKPVPKELISKP 101  
QY 108 --LGK-DHEILRRRIENGAKELWFFLOSELKKLKEGNEIQRHADEILLDGHHERSIM 164  
DB 102 SILGPVQVEQVRMLDRIEMFYLLHSQTIE-----NSTKILLET--QMISIM 148  
QY 165 TDLYLSQTDGAGEWREKADLTTELVRITTYLQNKDCSKARKLVNINKNGCGYGCOL 224  
DB 149 GLSAQLEKLEGESEERFPQRTAITQIRIPKSTKLNQPKACSEAKTLVNCNLKCEGFGCOL 208  
QY 225 HHVYCFMAYGTQRTIL--ESQNRVATGWETVFRPVSETCTDRSGLSTGHSWGE-- 280  
DB 209 HHVYCAITAFATQRMVVLKRGSSWKYSSHCWTSVFKLKSCEDE--AVGNTAEKPP 265  
QY 281 VKDKVQVVELPIVDSLHPRPYLPLAVPEDLADRLRVHGDPAVMWSQFVKYLIRPOP 340  
DB 266 AEPSPARVSLGIVDSLITKPTFLQAVPEQLLESILSLSHPPAFFVGTFTSYLMRFNS 325  
QY 341 WLERIEETTKLGF-KHPVIGVHVRTDKVCTEAPHPTEEMVHVEHFQLLERMK-K 398  
DB 326 ATQEKLDKALKSIPDKGPVIGLQIRRTDKVCTEAPHPTEEMVHVEHFQLLERMK 385

QY 399 VDKRVVYLATDDPSLLKEATKYSNYEFISDINSISWSAGLHNRNTENSLRGVILDIHFLS 458  
DB 386 PLERRIFIASDDPTVWPEAKNDYENYGVSTEIAKTAQLNRYTDSALMGVITDIYILS 445  
QY 459 QADELVCTFSQVCRVAYEIMQTLHPDASANFHSLLDIYFYGQNAHQAIVYHPQRTK 518  
DB 446 KNYLVCTFSQVCRMGVELRQPSGADGSKFHSLLDIYFYGQNAHQAIVYHPQRTK 505  
QY 519 BEIPMEPGDIIGVAGNHNWGSKGVNRKLGRTGLYPSYKVR 559  
DB 506 KEIDLKVGDKVGIAGNHNWGSKGTNRQTYKEGVFPYSKIR 546

RESULT 3  
T43527  
sp8 protein - fission yeast (Schizosaccharomyces pombe) (fragment)  
C:Species: Schizosaccharomyces pombe  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 04-Feb-2000  
C:Accession: T43527  
R:Jimenez, M.; Goday, C.  
submitted to the EMBL Data Library, June 1997  
A:Description: Schizosaccharomyces pombe myosin-like.  
A:Reference number: 222545  
A:Accession: T43527  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1173 <JIM>  
A:Cross-references: EMBL:AF010473; PIDN:AAB65416.1  
A:Experimental source: strain 972h(-)  
C:Genetics:  
A:Gene: sp8

Query Match 4.3%; Score 133; DB 2; Length 1173;  
Best Local Similarity 21.2%; Pred. No. 1.3;  
Matches 110; Conservative 74; Mismatches 226; Indels 108; Gaps 20;  
QY 34 DHPDHSRELSKILAKLERLKOQNEDLRRMAESLRIPEGPIDQGTATGRVRLBEQLVKA 93  
DB 612 NHVKRISKEMBIISQORQLLPLENKLRKTVESSNRVIADLQRG-----ITEKDVST 664  
QY 94 KEQI---ENY-----KKQARNDLGKDHEI---LRRRIENGAKELWFF----- 129  
DB 665 SESYGERSNLYNMVALLNESKLSRENLERNEEVIETREKIEITLKTDLANFLNKEOLE 724  
QY 130 --LOSELKKLKEGNE--ELQRHADEILLDGHHERSIMTDLYLSQTDGAGEWREKAK 185  
DB 725 SLOTEKAAVKKLENSNEEYKRNQEIILLSL--NSTSTSSDASRLK-----NELVSK-- 776  
QY 186 DLTELVRITTYLQNKDCSKARKLVNINKNGCGYGCOLHHVYCFMAYGTQRTILLES 245  
DB 777 NLIBELNQEIGHLKSELETVKSKSEDLNER-----AQNQSKIEQLEL 819  
QY 246 QNRVYATGWETVFRPVSETCTDRSGLSTGHSWGEVKDKNVQVVELPIVDSLHPRPPYLP 305  
DB 820 KNTKLA--AAMTKYQVQVNVKSLKHNQIRLOLSQKTSLEAKVAECHOQLNQLNPKSATP 878  
QY 306 LAVPEDLADRLRVHGDPAVMWSQFVKYLIRPOPWLERIEETTKKLGKHPVIGVHVR 365  
DB 879 TATTQ-----SEPTSVLSSEFNS-----TKEELSSTQKLSLMDILNTTKE 920  
QY 366 RTDKV-----GTEAAFH--PIBEYM-----VHVEHFQLLERMKVKKRVYLATDDPS 412  
DB 921 ELEKVRQNSKSEGTSDTEIPNEEEMERKVKMQQEVLRSLRSIAKELQKNELLRKQV 980  
QY 413 LLKEAK----TKYNYEFISDINSISWSAGLHNRNTENSLRGVILDIHF-----LSQA 460  
DB 981 LQDQVKALQETVWSSEEAESASVHADTKOLENLKKTTEEMLSVTFOVIFNESISDFSTSTA 1040  
QY 461 DFLVCTFSQVCRVAYEIMQTLHPDASANFH--SLDDI 496  
DB 1041 DF--TTFQKEWEKREILQKDVVEQVAQSHQKQLDNI 1076



RESULT 4  
T50073  
myosin-like coiled-coil protein sp8 [imported] - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C>Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 09-Jun-2000  
C:Accession: T50073  
R:McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Seeger, K.; Harris, D.  
submitted to the EMBL Data Library, December 1999  
A:Reference number: Z25034  
A:Accession: T50073  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1727 <MCD>  
A:Cross-references: EMBL:AL133357; PIDN: CAB62414.1; GSPDB: GN00066; SPDB: SPAC1486.04C  
A:Experimental source: strain 972h(-); cosmid c1486  
C:Genetics:  
A:Gene: SPDB: SPAC1486.04C  
A:Map position: 1

Query Match 4.3%; Score 133; DB 2; Length 1727;  
Best Local Similarity 21.2%; Pred. No. 2.3; Mismatches 226; Indels 108; Gaps 20;  
Matches 110; Conservative 74

Qy 34 DHPDSSRELKILAKLERLQKQEDLRMAESLRPEGPIDQGTATGRVRLVEQLVKA 93  
Db 1166 NAVKRIKEMELISQORQLLFLENKKLRTVSSNRVIADLQRG-----ITEKDVST 1218  
Qy 94 KEQI---ENY-----KKQANDLGKDHET---LRRRIENGAKELWFF----- 129  
Db 1219 SESVGERSNYLMVALLNESKSLRENLERNEEVITELREKTIETLKTDLANFRLNKEQLE 1278  
Qy 130 --LQSELKLLKLEGN--ELORHADEILLGHHERSIMTDLVYLSQTDGAGWEKREAK 185  
Db 1279 SLOQTEKAAVKKLENSNEYKRNHQQEILLSL-NSSTSTSSDASRLK-----NELVSK 1330  
Qy 186 DLTELVRRTVLPQKDCSKARKLVNINKGCGYGCQLHHVYCFMAYGTQRTLLLES 245  
Db 1331 NLIEELNOEIGHKSELTFTVSKSEDLNER-----AQNSKTEQLLEL 1373  
Qy 246 QNWRYATGQWETFRPVSECTTDRSGLSTGHWSEVGDKNVQVPELPIVDSLHPRPPYLP 305  
Db 1374 KNTKLA-AAWRTKYQVQVVKSLKHNQIRQLSQKTSLEAKVAECHOINQLNKPSTAP 1432  
Qy 306 LAVPEDLADRLLRVHGDPVAVWVVSQVKYLTPQPLWLEREEETTKLGFKPHGVHVR 365  
Db 1433 TATTO-----SEFSTVSLEEFNS-----TKEELSSQTKLSIMDLINTTKE 1474  
Qy 366 RTDKV-----GTEAAPH-PIEYV-----VHVEEHFOLLERRMKVKKRVYLATDDPS 412  
Db 1475 ELEKVRQNSKSEGTSTKTEIPNEEEMERKKVQVQEVLRSLRIAKELQKELLARKQNOV 1534  
Qy 413 LLKEAK---TKSYNYEFISDNSISWSAGLHNRNTENSLRGVILDIHF-----LSQA 460  
Db 1535 LQDQVKALQETVWSSEAEASVHADTKDLENLKKTEEMLSVTFQVFNESISDFSTSTA 1594  
Qy 461 DPLVCTFSQVCRVAYEIMQTLHPDASANFH--SLDDI 496  
Db 1595 DF--TTFVQKWEKREILQKDVQVQACSHQKQLDNI 1630

RESULT 5  
H88187  
protein C18H9.8 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
C:Accession: H88187  
R:Anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A:Reference number: A75000; MUID: 99069613; PMID: 9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/projects/C\_elegans/  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and

A:Accession: H88187  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-678 <STO>  
A:Cross-references: GB:chr\_XI; PID:g722387; GSPDB:GN00020; CBSP:C18H9.8  
C:Genetics:  
A:Gene: C18H9.8  
A:Map position: 2

Query Match 4.2%; Score 128; DB 2; Length 678;  
Best Local Similarity 19.5%; Pred. No. 1.4; Mismatches 82; Indels 108; Gaps 19;  
Matches 91; Conservative 82

Qy 38 HSSRELSKILAKLERLQKQEDLRMAESLRPEGPIDQGTATGRVRLVEQLVKAQKQI 97  
Db 255 HLNGDMSDIELEASKMKEQADEMSLAEV-----FNERRAKEELQGLETEV 302  
Qy 98 ENYK-----KQARNDLGKDHEITLRR--FENGAKELWFFIQLSELKKLK-KLE 141  
Db 303 EEQKLNLEAVTHAMPQMKKEYEDLKSEAKLLRRVVEAKN--EDLDDRISKEYEIR 360  
Qy 142 GNEIQRHA---DEILLGHHERSIMTDLVYLSQTDGAGWEKREAKDLTELVRRIYL 198  
Db 361 SNPLKKKAIQLQETLDTLKKOEKLMEDMQSALTPEA---WRDKMSENKMLNADLVVIE 417  
Qy 199 QNPK---DCSKARKLVNINKGCGYGCQLHHVYCFMAYGTQ-----RTLILES 245  
Db 418 KOHTVQDQISLASEELHEVDSQEAQIMAHHTKYLLLSKSTMLDDTTENYPOOVIYQ 477  
Qy 246 QNWRYATGQWETFRPVSE-----TCTDRSGLS---TGHWSGEVGDKNVQV--- 288  
Db 478 QDIEEFSDAVVLILRKISANLKVNLQEDITDLDERGLTLQTN-VDELKEMHVRLOBEL 536  
Qy 289 -----VELPI---VDSLHPRPPYLPVPELADRLLRVHGDPVAVWVVSQVKYLIRQP 340  
Db 537 ISIDMELALNEETDNLTEK-----KIDQELAGVGNVDS----- 574  
Qy 341 WLEREIETTKLGFKPHGVHVRRTDK--VGTEAAPHPIEYVHVEEHFOLLERRMK 398  
Db 575 GLRQELERQKRLDEAPARSHMQLEAVNASTNELHSLIPGYSQH-----KMLRERLE 629  
Qy 399 VDKKRVYLATDDPSLLKKEAKTKSYNYEFISDNSISWSAGLHNRNTEN 445  
Db 630 AVEKRTAAKSLDMSLRKTE---IDYEDIKTESIRLQBEYNNMLLTN 672

RESULT 6  
A33977  
myosin heavy chain, nonmuscle - chicken  
N:Contains: myosin ATPase (EC 3.6.4.1)  
C:Species: Gallus gallus (chicken)  
C>Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 19-Apr-2002  
C:Accession: A33977; S06116; A43422  
R:Shohet, R.V.; Conti, M.A.; Kawamoto, S.; Preston, Y.A.; Brill, D.A.; Adelstein, R.S.  
Proc. Natl. Acad. Sci. U.S.A. 86, 7726-7730, 1989  
A:Title: Cloning of the cDNA encoding the myosin heavy chain of a vertebrate cellular my  
A:Reference number: A33977; MUID: 90046668; PMID: 2813355  
A:Accession: A33977  
A:Molecule type: mRNA  
A:Residues: 1-1959 <SHO>  
A:Cross-references: GB:M26510; NID:g212382; PIDN:AAA48974.1; PID:g212383  
R:Katsuragawa, Y.; Yanagisawa, M.; Inoue, A.; Masaki, T.  
Eur. J. Biochem. 184, 611-616, 1989  
A:Title: Two distinct nonmuscle myosin-heavy-chain mRNAs are differentially expressed in  
s.  
A:Reference number: S06116; NUID: 90032648; PMID: 2806244  
A:Accession: S06116  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 716-1008 <KAT>  
A:Cross-references: GB:X17589  
A:Note: this translation is not annotated in GenBank entry GGMHCFMHA, release 114  
R:Hodge, T.P.; Cross, R.; Kendrick-Jones, J.

J. Cell Biol. 118, 1085-1095, 1992  
 A:Title: Role of the COOH-terminal nonhelical tailpiece in the assembly of a vertebrate  
 A:Reference number: A43422; MUID:92381096; PMID:1512291  
 A:Accession: A43422  
 A>Status: nucleic acid sequence not shown; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1900-1959 <HOD>  
 A:Experimental source: brush border  
 A>Note: sequence extracted from NCBI backbone (NCBIP:111947)  
 C:Superfamily: myosin heavy chain; myosin motor domain homology  
 C:Keywords: actin binding; ATP: coiled coil; hydrolase; methylated amino acid; nucleotide  
 F:84-764/Domain: myosin motor domain homology <MMOT>  
 F:174-181/Region: nucleotide-binding motif A (P-loop)  
 F:552-565/Region: actin binding #status predicted  
 F:626-640/Region: actin binding #status predicted  
 F:837-1936/Domain: coiled coil #status predicted <COI>  
 F:837-1277/Region: S2  
 F:1278-1959/Region: light meromyosin  
 F:1293-1959/Domain: carboxyl-terminal <CBT>  
 F:125/Modified site: N6,N6-trimethyllysine (Lys) #status predicted  
 F:180/Binding site: ATP (Lys) #status predicted  
 F:694,704/Active site: Cys #status predicted

Query Match 4.1%; Score 127.5; DB 1; Length 1959;  
 Best Local Similarity 19.8%; Pred. No. 6.1;  
 Matches 110; Conservative 79; Mismatches 177; Indels 189; Gaps 22;

QY 42 ELSKILAKLE-RLKQONEDLRMAESLRIPGPIQGTATGRVRLVLEQLVKAKEQIE-- 98  
 DB 1066 ELQQAELKILQSKKEELQ--ALARVEEBAQNNWALKRIELESQITELQEDLSE 1123  
 QY 99 ----NYKQARNDLGDHILRRRIEN-----GAKELWFFLQSELKLLKLEGNELQRH 148  
 DB 1124 RASRNKAERKQKDLGEELAKTELEDTLDSAAQELRSKEQEVTVLKKLEDEAKTH 1183  
 QY 149 ABEILLDGHHSIMTDLYL--SOTDGAWEKEKAOLTELVRRIITVQLNPQDCSKAR 208  
 DB 1184 EAQI-----QEMRQKHSQAIEELAEQ-----LEQTK 1209  
 QY 209 KLVCNKNGCGYCOLHHVVCYFMAYGTORTLLESQNRWATGWTETVPRVSETCTD 268  
 DB 1210 RVANLEKA-----KQALESER-----AELSN 1232  
 QY 269 RSGSLTGHWSGVEKQNV--QVVELPVDLSLHPRPPYPLAVPEDLADRL--LRVHGDP 324  
 DB 1233 VKVLLQKGDGAEHKKKVDQALQELQVKFTEGR-----VKTALAERNVKLQVELDNV 1285  
 QY 325 VNWVSQ----FVKYLIRPOPWLERI-----EETTKLGFKHPVIGVHVRRTDKVGT 372  
 DB 1286 TGLNQSDSKSIK-LAKDFALESALQDTQELLQEBETRLKLSF-----TKLKQT 1334  
 QY 373 EAAFPPIREYVHVEHFOLLERRMKVKKRYLATDPSLLKKAETKYSNVYFISDMSI 432  
 DB 1335 EDEKALKEQLEEBEAKNLEKQISVLQQAV-----EARKQMD----- 1374  
 QY 433 SWSAGLHNRYSNLSRGVLDIHFLSQADFLVCTFSQVCRVAYEIMQTLHPDASANPHS 492  
 DB 1375 -----DGLGC-----LETAEEAKKKLQKD-----LES 1396  
 QY 493 LDDIYFGQNAHNOIAVYPHOPRTKEIPMPGDIIGVAGNHWNYSKGVNKRKLGKTGL 552  
 DB 1397 LTQRY-----EETIAAYDKLEKTKRLRLQQLDLD-IAVDLHQRTVSNLEKKQK--- 1445  
 QY 553 YPSYKVRKEIETVKY 567  
 DB 1446 FQLLAEEKNISAKY 1460

## RESULT 7

S28061  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 08-Oct-1999

C:Accession: S28061  
 R:Meuwissen, R.L.J.; Offenberger, H.H.; Dietrich, A.J.J.; Riesewijk, A.; van Iersel, M.; H  
 ENBO J. 11, 5091-5100, 1992  
 A:Title: A coiled-coil related protein specific for synapsed regions of meiotic prophase  
 A:Reference number: S28061; MUID:93099884; PMID:1464329  
 A:Accession: S28061  
 A:Molecule type: mRNA  
 A:Residues: 1-946 <MEU>  
 A:Cross-references: EMBL:X67805; NID:957212; PIDN:CAA48006.1; PID:957213  
 C:Genetics:  
 A:Gene: SCPI  
 C:Keywords: DNA binding

Query Match 4.1%; Score 127; DB 2; Length 946;  
 Best Local Similarity 20.3%; Pred. No. 2.5;  
 Matches 105; Conservative 86; Mismatches 188; Indels 138; Gaps 23;

QY 42 ELSKILAKLERLQONEDLRMAESLRIPGPIQGTATGRVRLVLEQLVKAKEQIENY 100  
 DB 104 ENEKSVLKELEIEIQENKDL-----IKENNATRHWCNLLKKEKTCARSAEKTSKY 150  
 QY 101 K-----KQARNDLQKD-----HEILRRRIENGAKELWFFLQSELKLLKLEGNELQR 147  
 DB 151 EYERETROVYVDLNNIEKMLAFEBELRVQAEANRLEMHPKLEKHEKIQHLE-EEYQK 209  
 QY 148 HAD-----EILLDGHHSIMTDLYL--SOTDGAWEKEKAOLTELVRRIITVQL 199  
 DB 210 EVNNKENQVSLLLIQSTEKENKMDLTFLLSESRDKANQLEEK-----TKLQDENLKL 264  
 QY 200 NPKDCSARKLVNINCGYCOLHHVVCYFMAYGTORTLLESQNRWATGWTETV 259  
 DB 265 EKDHLS-----ELEDIKMSMQRSMTQKTL-----EEDLQIAT---KTIY 303  
 QY 260 RPVSE-----TCTDRSGLSLGHM----- 277  
 DB 304 QLTEKAQAEELNKAATHTSLVTELKATCTCTLEELLRTTEQOENEDOLKLITMELQ 363  
 QY 278 --SGEVDK---KNVQVVELPVDLSLHPRPPYPLAVPEDLADRLRVHG-DPAVWVVSQ 330  
 DB 364 KKSSELEEMTKFNKEVELEELKTI-----LAEDQKLLDEKKQVEKLAEBELQKEQ 415  
 QY 331 FVKYLI--RPOPWLERIEETTKLGFKHPVIGVHVRRTDKVGTAAAFHPITEEYVHVEE 388  
 DB 416 ELTELLQRTKEIHDLVEQVTVTKTSEH-----YLKQVEEMKTELEKEKLNIELTANS 470  
 QY 389 HFOLLERRMKVKKRYLATDPSLLKKAETKYSNVYFISDMSISWSAGLHNRYSNLSR 448  
 DB 471 DMLLE-----NKKLVQESDMVLELKKHQEDIIINCKQEBMLKQIETLEEK--ENNL 523  
 QY 449 GVILDI--HFLSQADFLVCTF--SSQVCR-VAYEIMQ 480  
 DB 524 DELESVRKEFIQGDVEVKCKLDKSEENARSIEYVLK 560

## RESULT 8

A61231  
 myosin heavy chain nonmuscle form A - human  
 N:Alternate names: cellular myosin heavy chain; myosin type 9; NMHC-A  
 N:Contains: myosin ATPase (EC 3.6.4.1)  
 C:Species: Homo sapiens (man)  
 C:Date: 12-May-1994 #sequence\_revision 14-Jul-1994 #text\_change 19-Apr-2002  
 C:Accession: A61231; A34876; I52562; I61692  
 R:Simons, M.; Wang, M.; McBride, O.W.; Kawamoto, S.; Yamakawa, K.; Gdula, D.; Adelstein,  
 Circ. Res. 69, 530-539, 1991  
 A:Title: Human nonmuscle myosin heavy chains are encoded by two genes located on different  
 A:Reference number: A61231; MUID:91316803; PMID:1860190  
 A:Accession: A61231  
 A:Molecule type: mRNA  
 A:Residues: 1-715 <SIM>  
 A:Cross-references: GB:M69180; NID:9189029; PIDN:AAA61765.1; PID:9189030  
 R:Saez, C.G.; Myers, J.C.; Shows, T.B.; Leinwand, L.A.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 1164-1168, 1990  
 A:Title: Human nonmuscle myosin heavy chain mRNA: generation of diversity through altern







```
Db 500 RUT-----TORHEI-ESTN----- 512
QY 273 STGWSGEVKDKNVQVLP-IVDSLHPRPPVLPPLAVPE--DLADRLRV-----HGDDPA 324
Db 513 -----KSELRIAEITHLQOQLQESQQLGKMIPEKQSLIDQLQVQNSLHRDSL 563
QY 325 VVWSQFVKYLRPFWLREIETTKLGFHPVIGVHVR-RTDKV--GTEAAPHPIEE 381
Db 564 L-----TLKRALE--TKEIGRQ-----QLRQLDEVEKETRAKLQEIDV 600
QY 382 YNVVHEHFQLLERRM-----KVDKKVYLATDDPSLLKEAKTKSYNYEISDNS 431
Db 601 FNNQLKELRELNYKQOFQKQDFETEKIKQELERKTSSELDKLBKEDKR-----RMLQDQK 656
QY 432 ISWSAGLHNRVTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQTLHPDASANFH 491
Db 657 L-W-----QDRVKQBEERYKFQDEEKEREE-----SVQKCEV--EKKPEIQEKNKPPH 703
QY 492 SLDDIYFQGG-----NAH-----NO-----IAYVPHQPRTKKEEIPMBPGDIIGVAG 533
Db 704 QPPEFGKLGQIPWMTTEKAPLTINQGVVYRYALYPFDARSHDEITIEPBGDIIMVDE 763
QY 534 NH-----WNGYSKGVNRKLGKGLYPSYKREKIETVKYP--TYPEAE 574
Db 764 SQTGFGMLG-----GELKGTGWFFA-NYAERMPSEFPSTTKPAE 805

RESULT 15
B43402
myosin heavy chain-B, neuronal - chicken
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Gallus gallus (chicken)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C:Accession: B43402; A43402
J:Takahashi, M.; Kawamoto, S.; Adelstein, R.S.
R. Biol. Chem. 267, 17864-17871, 1992
A:Title: Evidence for inserted sequences in the head region of nonmuscle myosin specific
yosin.
A:Reference number: A43402; MUID:92388144; PMID:1355479
A:Accession: B43402
A:Molecule type: mRNA
A:Residues: 1-2007 <TAK>
A:Cross-references: GB:M93676; NID:G212448; PIDN:AAA48988.1; PID:G212452
A>Note: the sequence of residues 212-221 and 632-652 and the corresponding nucleotide se
A:Accession: A43402
A:Molecule type: mRNA
A:Residues: 1-211;222-631;653-2007 <TA2>
A:Cross-references: GB:M93676; NID:G212448; PIDN:AAA48985.1; PID:G212449
A>Note: sequence extracted from NCBI backbone (NCBIN:112864)
C:Comment: Alternatively spliced segments 1 and 2 are found exclusively in nonmuscle myo
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methylated
F:1-2007/Product: myosin heavy chain-B, neuronal #status predicted <MYN>
F:1-211;222-631;653-2007/Product: myosin heavy chain-B, nonmuscle #status predicted <MYS
F:88-802/Domain: myosin motor domain homology <MMOT>
F:178-185/Region: nucleotide-binding motif A (P-loop)
F:212-221/Region: alternatively spliced segment 1 #status experimental
F:559-593/Region: actin binding #status predicted
F:632-652/Region: alternatively spliced segment 2 #status experimental
F:692-714/Region: actin binding #status predicted
F:875-2007/Domain: coiled coil #status predicted <COI>
F:875-1315/Region: S2
F:1316-2007/Region: light meromyosin
F:129/Modified site: N6, N6-trimethyllysine (Lys) #status predicted
F:184/Binding site: ATP (Lys) #status predicted
F:732,742/Active site: Cys #status predicted
F:1954/Binding site: phosphate (Thr) (covalent) #status predicted
F:1987/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 3.9%; Score 119.5; DB 1; Length 2007;
Best Local Similarity 24.2%; Pred. NO. 21;
Matches 44; Conservative 37; Mismatches 46; Indels 55; Gaps 8;
```

```
QY 40 SRELSKILAKLERLKQONEDLRMAESLRIPGPIDQGTATGRVVRVLEQLVKAKEQIEN 99
Db 1359 TRQKLNLSRIRQLEEEKNLQEQE-----EEEEAR---KNLEKQMLAQOQLAE 1406
QY 100 YKQARNDLG-----KDHEILRRRIENGAKELWFFLQSELKKLKGEGNELQ 146
Db 1407 AKKVDDDDGLTIEGLEENKKLLKDMESLSORLEKA-----MAYDKLEKTK- ---NRLQ 1457
QY 147 RHADEIILLDLGHHSIMTDL-----YYLSQTDGA-GEWREKEAKD 186
Db 1458 QELDDLWDL-DHQRIVSNLEKKQKFFDQMLAEKKNISARYAEERDRAEAAREKETKA 1516
QY 187 LT 188
Db 1517 LS 1518

Search completed: February 2, 2004, 08:43:55
Job time : 21 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 08:39:39 ; Search time 10.5 Seconds  
(without alignments)  
2575.272 Million cell updates/sec

Title: US-09-971-773-23  
Perfect score: 3081  
Sequence: 1 MRAWTGSWRIMLILFAWTG.....YKVRKIETVKYTPAEAK 575

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 segs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3013	97.8	575	1	FUT8_MOUSE
2	2984	96.9	575	1	FUT8_MOUSE
3	2939	95.4	575	1	FUT8_BOVIN
4	2885	93.6	575	1	FUT8_PIG
5	133	4.3	1727	1	ALM1_SCHPO
6	130.5	4.2	751	1	COG5_DROME
7	127.5	4.1	1959	1	MYH9_CHICK
8	127	4.1	997	1	SCP1_RAT
9	124.5	4.0	845	1	SCP1_MESAU
10	124.5	4.0	1960	1	MYH9_HUMAN
11	124.5	4.0	1976	1	MYHA_BOVIN
12	120.5	3.9	978	1	RA50_AQUAE
13	119.5	3.9	1270	1	ITN1_XENLA
14	118.5	3.8	976	1	SCP1_HUMAN
15	118.5	3.8	1427	1	REST_HUMAN
16	118	3.8	1217	1	ITN1_RAT
17	116.5	3.8	1961	1	MYH9_RAT
18	116	3.8	214	1	OSF1_HUMAN
19	116	3.8	993	1	SCP1_MOUSE
20	116	3.8	2442	1	CEP2_HUMAN
21	115.5	3.7	3680	1	DMD_CANFA
22	115	3.7	896	1	RA50_THEAC
23	115	3.7	1027	1	KINN_MOUSE
24	115	3.7	1976	1	MYHA_HUMAN
25	115	3.7	3210	1	CENF_HUMAN
26	114.5	3.7	2017	1	MYGN_DROME
27	114	3.7	1976	1	MYHA_RAT
28	112.5	3.7	1935	1	MYSS_CYPCA
29	112.5	3.7	1938	1	MYSS_CHICK
30	112	3.6	2415	1	SPCA_MOUSE
31	111.5	3.6	880	1	RA50_PYRAB
32	111.5	3.6	919	1	RA50_AERPE
33	111.5	3.6	2230	1	GOG4_HUMAN

#### RESULT 1

ID	FUT8_MOUSE	STANDARD;	PRT;	575 AA.
AC	O9WT52; O921U1;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Alpha-(1,6)-fucosyltransferase (EC 2.4.1.68) (Glycoprotein 6-alpha-L-fucosyltransferase)			
DE	(GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alpha1,6-fucosyltransferase)			
DE	(alpha1-6FucT) (Fucosyltransferase 8)			
GN	FUT8.			
OS	Mus musculus (Mouse)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euteheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=20358720; PubMed=10902914;			
RA	Hayashi H., Yoneda A., Asada M., Ikekita M., Imamura T.;			
RT	"Molecular cloning of mouse alpha-1,6-fucosyltransferase and expression of its mRNA in the developing cerebrum.";			
RL	DNA Seq. 11:91-96 (2000).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE=2388257; PubMed=12477932;			
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,			
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,			
RA	Bosak S.A., McWeney P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,			
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Fabey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,			
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Rodriguez A.C., Richmond J., Schmutz J., Myers R.M.,			
RA	Butterfield Y.S.N., Krywinski M.I., Skalska U., Skalska D.E.,			
RA	Butcher J.A., Schein J.E., Jones S.J.M., Marra M.A.;			
RT	"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).			
CC	FUNCTION: Catalyzes the addition of fucose in alpha 1-6 linkage to the first GlcNAc residue, next to the peptide chains in N-glycans (By similarity).			
CC	CAATALYTIC ACTIVITY: GDP-L-fucose + N(4)-[N-acetyl-beta-D-glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)]-[N-acetyl-beta-D->4]-N-acetyl-beta-D-glucosaminyl-(1->4)-N-acetyl-beta-D-glucosaminyl]asparagine = GDP + N(4)-[N-acetyl-beta-D-glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)]-[N-acetyl-beta-D-			

#### ALIGNMENTS

34	111	3.6	933	1	PLD2_HUMAN
35	111	3.6	933	1	PLD2_RAT
36	111	3.6	1549	1	SMC4_CABEL
37	111	3.6	1597	1	CTRO_MOUSE
38	111	3.6	1938	1	MYHD_HUMAN
39	110.5	3.6	1969	1	MYSA_CABEL
40	110.5	3.6	2349	1	TPR_HUMAN
41	110	3.6	366	1	RP2_BACSU
42	110	3.6	558	1	YH32_AQUAE
43	110	3.6	1939	1	MYH4_HUMAN
44	109.5	3.6	579	1	GL60_HUMAN
45	109.5	3.6	1114	1	RH18_YEAST

O14939 homo sapien  
P70498 rattus norv  
Q20060 caenorhabdi  
P49025 mus muscucu  
Q9ukx3 homo sapien  
P12844 caenorhabdi  
P28367 bacillus su  
O67622 aquifex aeo  
Q9v623 homo sapien  
Q08378 homo sapien  
Q12749 saccharomyc



CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1->6)]-  
CC >4)-N-acetyl-beta-D-glucosaminyl-(1->4)-[alpha-L-fucosyl-(1->6)]-  
CC N-acetyl-beta-D-glucosaminyl]asparagine.  
CC -!- PATHWAY: Glycosylation.  
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound  
CC form in trans cisternae of Golgi (By similarity).  
CC -!- SIMILARITY: Contains 1 SH3 domain.  
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 23.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AB025198; BAA76392.1; -;  
CC EMBL; BC010666; AAH10666.1; -;  
CC MGD; MGI:1858901; Fut8.  
CC InterPro; IPR001452; SH3.  
CC Pfam; PF00018; SH3; 1.  
CC SMART; SM00326; SH3; 1.  
CC PROSITE; PS50002; SH3; FALSE NEG.  
CC Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;  
CC Golgi stack; SH3 domain.  
CC DOMAIN 1 9  
CC TRANSMEM 10 30 CYTOPLASMIC (POTENTIAL).  
CC SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
CC (POTENTIAL).  
CC LUMENAL, CATALYTIC (POTENTIAL).  
CC DOMAIN 31 575  
CC SITE 502 563 SH3.  
CC SITE 299 305 SH3-BINDING (POTENTIAL).  
CC SITE 365 366 IMPORTANT FOR DONOR-SUBSTRATE BINDING.  
CC FT CONFLICT 40 40 T -> S (IN REF. 2).  
CC FT CONFLICT 388 388 Q -> E (IN REF. 2).  
CC FT CONFLICT 418 418 N -> K (IN REF. 2).  
CC FT CONFLICT 418 418 N -> K (IN REF. 2).  
CC SEQUENCE 575 AA; 66555 MW; 7BE2ED3146E0C45F CRC64;  
Query Match 97.8%; Score 3013; DB 1; Length 575;  
Best Local Similarity 97.0%; Pred. No. 1.9e-193;  
Matches 558; Conservative 11; Mismatches 6; Indels 0; Gaps 0;  
QY 1 MRAWTGSRWIMLILFAWGTLLFYGGHLVRNDHPDHSRSLSKILAKLERLQONEDL 60  
DB 1 MRAWTGSRWIMLILFAWGTLLFYGGHLVRNDHPDHSRSLSKILAKLERLQONEDL 60  
QY 61 RMAESLRIPGPIDGTATGRVLEEQLVKAKEQIENYKQARNGLKQHEILRRRIE 120  
DB 61 RMAESLRIPGPIDGTATGRVLEEQLVKAKEQIENYKQARNGLKQHEILRRRIE 120  
QY 121 NGAKELWFFLQSELKKLEGNELQRADEILDLGHHERSIMTDLVYLSOTDCAGEWR 180  
DB 121 NGAKELWFFLQSELKKLEGNELQRADEILDLGHHERSIMTDLVYLSOTDCAGEWR 180  
QY 181 EKEADLTQLVQRRITLQNPDCSKARKLVNKNKGGYQQLHHVYCFMAYGTQRT 240  
DB 181 EKEADLTQLVQRRITLQNPDCSKARKLVNKNKGGYQQLHHVYCFMAYGTQRT 240  
QY 241 LILESQNRWYATGHWETVRPVSCTDRSGLSGTGHSGEVKQKVQVVELPVDLSLHPR 300  
DB 241 LILESQNRWYATGHWETVRPVSCTDRSGLSGTGHSGEVKQKVQVVELPVDLSLHPR 300  
QY 301 PPYLPLAVPEDLADLLRVLHGPDPVWWSQFVKYLRPOPMLERIEETTKLGFKHPVI 360  
DB 301 PPYLPLAVPEDLADLLRVLHGPDPVWWSQFVKYLRPOPMLERIEETTKLGFKHPVI 360  
QY 361 GVHVRRTDKVGTAAFPPIEEYVWVHEHFQLLERMKVDRKRVYLATDDPSLLKAKTK 420  
DB 361 GVHVRRTDKVGTAAFPPIEEYVWVHEHFQLLERMKVDRKRVYLATDDPSLLKAKTK 420  
QY 421 YSNYEFISDNTSWSAGLHNRYTENSRLRGVILDIHFLSQADFLVCTFSSQVCRVAYEMQ 480  
DB 421 YSNYEFISDNTSWSAGLHNRYTENSRLRGVILDIHFLSQADFLVCTFSSQVCRVAYEMQ 480

QY 481 TLHPDASANFSLDDIYFGQNANQIAVTPHQRTKEEIPMEPGDIIGVAGNHWGYS 540  
DB 481 TLHPDASANFSLDDIYFGQNANQIAVTPHQRTKEEIPMEPGDIIGVAGNHWGYS 540  
QY 541 KGVNRKLGKTKGLYPSYKVKREKIVKYPTYPEAEK 575  
DB 541 KGVNRKLGKTKGLYPSYKVKREKIVKYPTYPEAEK 575

## RESULT 2

FUT8\_HUMAN  
ID FUT8\_HUMAN STANDARD; PRT; 575 AA.  
AC Q9BYC5; O00235; Q9BYC6; Q9P2U5; Q9P2U6;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Alpha-(1,6)-fucosyltransferase (EC 2.4.1.68) (Glycoprotein 6-alpha-L-  
DE fucosyltransferase) (GDP-fucose-glycoprotein fucosyltransferase)  
DE (GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alpha1,6-fucosyltransferase)  
DE (alpha1-6FucT) (Fucosyltransferase 8).  
GN FUT8.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1), AND FUNCTION.  
EX MEDLINE=97279058; PubMed=9133635;  
RA Taniguchi N.;  
RA Yanagidani S.; Uozumi N.; Ihara Y.; Miyoshi E.; Yamaguchi N.;  
RT "Purification and cDNA cloning of GDP-L-Fuc:N-acetyl-beta-D-  
RT glucosaminide:alpha1-6 fucosyltransferase (alpha1-6 FucT) from human  
RT gastric cancer MKN45 cells.";  
RL J. Biochem. 121:626-632(1997).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Embryo;  
RT "Differential splice variants of human FUT8 embryonic cDNA.";  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBAJ databases.  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RX MEDLINE=20275614; PubMed=10814706;  
RA Yanaguchi Y.; Ikeda Y.; Takahashi T.; Ihara H.; Tanaka T.; Sasho C.;  
RA Uozumi N.; Yanagidani S.; Inoue S.; Fujii J.; Taniguchi N.;  
RT "Genomic structure and promoter analysis of the human  
RT alpha1,6-fucosyltransferase gene (FUT8).";  
RL Glycobiology 10:637-643(2000).  
RN [4]  
RP DONOR SUBSTRATE BINDING, AND MUTAGENESIS OF ARG-365 AND ARG-366.  
RX MEDLINE=20229550; PubMed=10764839;  
RA Takahashi T.; Ikeda Y.; Tateishi A.; Yamaguchi Y.; Ishikawa M.;  
RA Taniguchi N.;  
RT "A sequence motif involved in the donor substrate binding by  
RT alpha1,6-fucosyltransferase: the role of the conserved arginine  
RT residues.";  
RL Glycobiology 10:503-510(2000).  
CC -!- FUNCTION: Catalyzes the addition of fucose in alpha 1-6 linkage to  
CC the first GlcNAc residue, next to the peptide chains in N-glycans.  
CC -!- CATALYTIC ACTIVITY: GDP-L-fucose + N(4)-[N-acetyl-beta-D-  
CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)]-N-acetyl-beta-D-  
CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1-  
CC >4)-N-acetyl-beta-D-glucosaminyl-(1->4)-N-acetyl-beta-D-  
CC glucosaminyl]asparagine = GDP + N(4)-[N-acetyl-beta-D-  
CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)]-N-acetyl-beta-D-  
CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1-  
CC >4)-N-acetyl-beta-D-glucosaminyl-(1->4)-[alpha-L-fucosyl-(1-  
CC >6)]-N-acetyl-beta-D-glucosaminyl]asparagine.  
CC -!- PATHWAY: Glycosylation.  
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound  
CC form in trans cisternae of Golgi (By similarity).  
CC -!- ALTERNATIVE PRODUCTS:



```
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=QB9YC5-1; Sequence-Displayed;
CC Name=2; Synonyms=Retinal;
CC IsoId=QB9YC5-2; Sequence-VSP 001807, VSP_001808;
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 23.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; D89289; BAA19764.1; -
DR EMBL; AF038281; AAB92372.2; -
DR EMBL; AF038280; AAB92372.2; JOINED.
DR EMBL; Y17979; CAA76988.1; -
DR EMBL; Y17976; CAA76985.1; -
DR EMBL; Y17977; CAA76986.1; -
DR EMBL; Y17978; CAA76987.1; -
DR EMBL; AB049828; BAB40975.1; -
DR EMBL; AB049740; BAB40929.2; -
DR EMBL; AB032573; BAA92859.2; -
DR EMBL; AB032572; BAA92858.1; -
DR EMBL; AB032568; BAA92858.1; JOINED.
DR EMBL; AB032569; BAA92858.1; JOINED.
DR EMBL; AB032570; BAA92858.1; JOINED.
DR EMBL; AB032571; BAA92858.1; JOINED.
DR PIR; JC5432; JC5432.
DR Genew; HGNC:4019; FUT8.
DR MIM; 602589; -
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS0002; SH3; FALSE NEG.
KW Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;
KW Golgi stack; SH3 domain; Alternative splicing.
FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 10 30 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT LUMENAL; CATALYTIC (POTENTIAL).
FT SH3.
FT SH3-BINDING (POTENTIAL).
FT IMPORTANT FOR DONOR SUBSTRATE BINDING.
FT EVKDNQVQVVELPIVDSLPRPPYPLAVPEDLADRLVRVH
FT GDPVAVWVS -> TPIMLLVITLPPGQDCTIDTQKIHV
FT E (in isoform 2).
FT /FTid=VSP 001807.
FT Missing (in isoform 2).
FT /FTid=VSP 001808.
FT R->A,K: COMPLETE LOSS OF ACTIVITY.
FT R->A,K: DECREASES ACTIVITY TO 3%.
SQ SEQUENCE 575 AA; 66515 MW; 5AE24A93881E18D0 CRC64;
Query Match 96.9%; Score 2984; DB 1; Length 575;
Best Local Similarity 96.2%; Pred. No. 1.7e-191;
Matches 553; Conservative 10; Mismatches 12; Indels 0; Gaps 0;
Qy 1 MRAWTGSRWIMLILFANGTLLFYIGGHLVRDNDHPDHSRSLKILAKLRKQONEDL 60
Db 1 MRPTWGSRWIMLILFANGTLLFYIGGHLVRDNDHPDHSRSLKILAKLRKQONEDL 60
Qy 61 RMMAESLRIPGEPIDQGTATGRVRVLEQLVKAKQIENYKQARNDLGKDHILRRRIE 120
Db 61 RMMAESLRIPGEPIDQGTATGRVRVLEQLVKAKQIENYKQARNDLGKDHILRRRIE 120
Qy 121 NGAKELWFFLQSELKKLKEGNELEQRUADEILLDLGHHERSIMTDLYLVSQTDGAGWR 180
Db 121 NGAKELWFFLQSELKKLKEGNELEQRUADEILLDLGHHERSIMTDLYLVSQTDGAGWR 180
```

```
Qy 181 EKEADLTTELVRRTITLQNPDKCSKARKLVCNINKGCGYCOLHHVYVCFMAYGTORT 240
Db 181 EKEADLTTELVRRTITLQNPDKCSKARKLVCNINKGCGYCOLHHVYVCFMAYGTORT 240
Qy 241 LILESQNRWYATGWTETFRVSETCTDRSLSTGHWSGEVKNVQVVELPIVDSLHPR 300
Db 241 LILESQNRWYATGWTETFRVSETCTDRSLSTGHWSGEVKNVQVVELPIVDSLHPR 300
Qy 301 PPYLPLAVPEADLRLVRHGDPAVWVSQVVKYLIRPQWLEIREIETTKLGFKHPVI 360
Db 301 PPYLPLAVPEADLRLVRHGDPAVWVSQVVKYLIRPQWLEIREIETTKLGFKHPVI 360
Qy 361 GVHVRRTDKVTEAAPHPIEYVMVHVEHFFOLLERRMKVKRKYVLATDDPSLLKEATK 420
Db 361 GVHVRRTDKVTEAAPHPIEYVMVHVEHFFOLLARRMQVKKRKYVLATDDPSLLKEATK 420
Qy 421 YSNYEFISDINSISWSAGLHNRNTENSRLGVLDLHFLSQADFLVCTFSSQVCRVAYEIMQ 480
Db 421 YPNYEFISDINSISWSAGLHNRNTENSRLGVLDLHFLSQADFLVCTFSSQVCRVAYEIMQ 480
Qy 481 TLHPDASANFHSLLDIYFVGQNAHNOIAVYPHOPRTKEEIPMPBPDIIGVAGNHWGYS 540
Db 481 TLHPDASANFHSLLDIYFVGQNAHNOIAVYAHQPTADEIPMPBPDIIGVAGNHWGYS 540
Qy 541 KGVNRKLGKTGLPYPSYKVKREKIETVKYPTYPEAEK 575
Db 541 KGVNRKLGRTGLPYPSYKVKREKIETVKYPTYPEAEK 575
RESULT 3
FUT8_BOVIN
ID FUT8_BOVIN STANDARD; PRT; 575 AA.
AC Q9N0W2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Alpha-(1,6)-fucosyltransferase (EC 2.4.1.68) (Glycoprotein 6-alpha-L-
DE fucosyltransferase) (GDP-fucose--glycoprotein fucosyltransferase)
DE (GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alpha1,6-fucosyltransferase)
DE (alpha1-6FucT) (Fucosyltransferase 8).
GN Fut8.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=20523969; PubMed=11070054;
RA Javard C., Dupuy F., Maftah A., Michalski J.-C., Oriol R.,
RA Petit J.-M., Julien R.;
RT "Ancestral exonic organization of FUT8, the gene encoding the
RT alpha6-fucosyltransferase, reveals successive peptide domains which
RT suggest a particular three-dimensional core structure for the
RT alpha6-fucosyltransferase family.";
RL Mol. Biol. Evol. 17:1651-1672(2000).
CC -!- FUNCTION: Catalyzes the addition of fucose in alpha 1-6 linkage to
CC the first GlcNAc residue, next to the peptide chains in N-glycans
CC (By similarity).
CC -!- CATALYTIC ACTIVITY: GDP-L-fucose + N(4)-{N-acetyl-beta-D-
CC Glucosaminyl- (1->2)-alpha-D-mannosyl- (1->3)-[N-acetyl-beta-D-
CC Glucosaminyl- (1->2)-alpha-D-mannosyl- (1->6)]-beta-D-mannosyl- (1-
CC >4)-N-acetyl-beta-D-glucosaminyl- (1->4)-N-acetyl-beta-D-
CC Glucosaminyl]asparagine = GDP + N(4)-{N-acetyl-beta-D-
CC Glucosaminyl- (1->2)-alpha-D-mannosyl- (1->3)-[N-acetyl-beta-D-
CC Glucosaminyl- (1->2)-alpha-D-mannosyl- (1->6)]-beta-D-mannosyl- (1-
CC >4)-N-acetyl-beta-D-glucosaminyl- (1->4)-[alpha-L-fucosyl- (1->6)]-
CC N-acetyl-beta-D-glucosaminyl]asparagine.
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound
CC form in trans cisternae of Golgi (By similarity).
CC -!- TISSUE SPECIFICITY: Highest expression found in brain. Also found
```

in heart, lung, spleen and kidney.  
 -!- SIMILARITY: Contains 1 SH3 domain.  
 -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 23.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL: AF247186; AAF65460.1; --  
 InterPro: IPR001452; SH3.  
 Pfam: PF00018; SH3; 1.  
 SMART: SM00326; SH3; 1.  
 PROSITE: PS00002; SH3; 1.  
 Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;  
 KW Golgi stack; SH3 domain.  
 FT DOMAIN 1 9  
 FT TRANSMEM 10 30  
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT LUMENAL, CATALYTIC (POTENTIAL).  
 FT SH3.  
 FT SH3-BINDING (POTENTIAL).  
 FT IMPORTANT FOR DONOR SUBSTRATE BINDING (BY  
 FT SIMILARITY).  
 FT  
 SQ SEQUENCE 575 AA; 66492 MW; FE04C66B8A5BF540 CRC64;  
 -----  
 Query Match 95.4%; Score 2939; DB 1; Length 575;  
 Best Local Similarity 94.4%; Pred. No. 1.6e-188;  
 Matches 543; Conservative 15; Mismatches 17; Indels 0; Gaps 0;  
 -----  
 QY 1 MRATGSGWRWIMLLIFAWGTLFLFYGGHLVRNDHPDHSRSLSKILAKLERLKKQNEDL 60  
 Db 1 MRPTGSGWRWIMLLIFAWGTLFLFYGGHLVRNDHPDHSRSLSKILAKLERLKKQNEDL 60  
 QY 61 RMAESLRIPGPIDQGTATGVRVRVLEQLVKAKEQIENYKQARNDLGKHEILRRRIE 120  
 Db 61 RMAESLRIPGPIDQGPASGRIRALEQLVKAKEQIENYKQTRNGLGKDEILRRRIE 120  
 QY 121 NGAKELWFLQSELKLLKLEGNELQRADEILDLGHHSRIMTDLYLSQTDGAGWR 180  
 Db 121 NGAKELWFLQSELKLLKLEGNELQRADEFLSLGHHSRIMTDLYLSQTDGAGWR 180  
 QY 181 EKEAKDLTVLORRITYLQNPDCSKARKLVNCKGCGQLHHVVCYCFMAYGTQT 240  
 Db 181 EKEAKDLTVLORRITYLQNPDCSKARKLVNCKGCGQLHHVVCYCFMAYGTQT 240  
 QY 241 LILESQNRVYATGGWETVRPVSETCTDRSGLSTGHSGEVKDKVQVVELPVDLSLHPR 300  
 Db 241 LILESHNRVYATGGWETVRPVSETCTDRSGVYTGHSGEIKDKNVQVVELPVDLSLHPR 300  
 QY 301 PLYLPLAVPEDLADLLRVHGDPAVWVSQFYKYLIRPQWLERIEETTKLGFKHPVI 360  
 Db 301 PLYLPLAVPEDLADLLRVHGDPAVWVSQFYKYLIRPQWLERIEETTKLGFKHPVI 360  
 QY 361 GVHVRTKVGTEAAPHFTEEVHVEHFQLLERRMKVVKRVYLATDDPSLLKEATK 420  
 Db 361 GVHVRTKVGTEAAPHFTEEVHVEHFQLLARRMQVKRVYLATDDPSLLKEATK 420  
 QY 421 YSNYEFISDINSISWSAGLHNRVTENSLRGVILDIHFLSQADFLVCTFFSQVCRVAYEIMQ 480  
 Db 421 YPHYEFISDINSISWSAGLHNRVTENSLRGVILDIHFLSQADFLVCTFFSQVCRVAYEIMQ 480  
 QY 481 TLHPDASANFSLDDIYFPGQNAHQVAYPHQRTKEEIPMEPCDIIGVAGNHWGYS 540  
 Db 481 TLHPDASANFSLDDIYFPGQNAHQVAYPHPEPTADEIPMEPCDIIGVAGNHWGYS 540  
 QY 541 KGVNRKLGKTLGYPYKVKREKETVKYPTYPEAK 575  
 Db 541 KGVNRKLGKTLGYPYKVKREKETVKVPHVPEAK 575

## RESULT 4

FUT8\_PIG  
 ID FUT8\_PIG STANDARD; PRT; 575 AA.  
 AC P79282;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Alpha-(1,6)-fucosyltransferase (EC 2.4.1.68) (Glycoprotein 6-alpha-L-fucosyltransferase) (GDP-fucose--glycoprotein fucosyltransferase)  
 DE (GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alpha1,6-fucosyltransferase)  
 DE (alpha1-6fuct) (Fucosyltransferase 8).  
 GN FUT8  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OC NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A., SEQUENCE OF 102-130; 333-344; 402-415 AND 566-575, AND FUNCTION.  
 RC TISSUE=Brain;  
 RX MEDLINE=97066976; PubMed=8910378;  
 RA Uozumi N., Yanagidani S., Miyoshi E., Ihara Y., Sakuma T., Gao C.-X., Teshima T., Fujii S., Shiba T., Taniguchi N.;  
 RT "Purification and cDNA cloning of porcine brain GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alpha1--6fucosyltransferase.";  
 RL J. Biol. Chem. 271:27810-27817(1996).  
 CC -!- FUNCTION: Catalyzes the addition of fucose in alpha 1-6 linkage to the first GlcNAc residue, next to the peptide chains in N-glycans. The pH optimum is 7.  
 CC -!- CATALYTIC ACTIVITY: GDP-L-fucose + N(4)-{N-acetyl-beta-D-glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)-[N-acetyl-beta-D-glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1->4)-N-acetyl-beta-D-glucosaminyl-(1->4)-N-acetyl-beta-D-glucosaminyl}asparagine = GDP + N(4)-{N-acetyl-beta-D-glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)-[N-acetyl-beta-D-glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1->4)-N-acetyl-beta-D-glucosaminyl}asparagine.  
 CC -!- PATHWAY: Glycosylation.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound form in trans cisternae of Golgi (By similarity).  
 CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN BRAIN.  
 CC -!- SIMILARITY: Contains 1 SH3 domain.  
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 23.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL: D86723; BAA13157.1; --  
 InterPro: IPR001452; SH3.  
 Pfam: PF00018; SH3; 1.  
 SMART: SM00326; SH3; 1.  
 PROSITE: PS00002; SH3; 1.  
 Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;  
 KW Golgi stack; SH3 domain.  
 FT DOMAIN 1 9  
 FT TRANSMEM 10 30  
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT LUMENAL, CATALYTIC (POTENTIAL).  
 FT SH3.  
 FT SH3-BINDING (POTENTIAL).  
 FT IMPORTANT FOR DONOR SUBSTRATE BINDING (BY  
 FT SIMILARITY).  
 FT  
 SQ SEQUENCE 575 AA; 66229 MW; 0F199D0BC2018F7B CRC64;  
 -----  
 Query Match 93.6%; Score 2885; DB 1; Length 575;  
 Best Local Similarity 92.9%; Pred. No. 6.5e-185;  
 -----

```
Matches 534; Conservative 16; Mismatches 25; Indels 0; Gaps 0;
QY 1 MRWGTGSRWMLILFANGTLLFYIGGHLVRDNDHPDSSRELKILAKLRLKQONEDL 60
DB 1 MRPTGSRWMLILFANGTLLFYIGGHLVRDNDHSDHSSRELKILAKLRLKQONEDL 60
QY 61 RMAESLRIPGPDIDQGTATGRVRLBQLVKAKEQIENYKQARNDLKGKDHIELRRRIE 120
DB 61 RMAESLRIPGPDIDQGTATGRVRLBQLVKAKEQIENYKQARNDLKGKDHIELRRRIE 120
QY 121 NGAKELWFFLOSELKLLKLEGNELQRADEFLDLGHHERSIMTDLYLSQTDGAGBWR 180
DB 121 NGAKELWFFLOSELKLLKLEGNELQRADEFLDLGHHERSIMTDLYLSQTDGAGBWR 180
QY 181 EKEAKDLTELVRRTIYLNPKDCSKARKLVNINKGCGYCOLHHVYVCFMIAVGTORT 240
DB 181 EKEAKDLTELVRRTIYLNPKDCSKARKLVNINKGCGYCOLHHVYVCFMIAVGTORT 240
QY 241 LILESQNRWYATGWTETFRPVSETCTDRSGLSTGHSGEVKDKNVQVVELPIVDSLHPR 300
DB 241 LALESHNRWYATGWTETFRPVSETCTDRSGSSTGHSGEVKDKNVQVVELPIVDSVHPR 300
QY 301 PPYLPLAVPEDLADRLRLRHGDPAAWVVSQFVKYLIRPQWLEKEIEEATKGLGKHPVI 360
DB 301 PPYLPLAVPEDLADRLVRHGDPAWVVSQFVKYLIRPQWLEKEIEEATKGLGKHPVI 360
QY 361 GVHVVRTDKVTEAAPHPIEEMVHVEHFOLLERMKVKKRVYLATDDPSLLKEATK 420
DB 361 GVHVVRTDKVGAAPHPIEETVHVEDFDQLARRMQVKKRVYLATDDPSLLKEATK 420
QY 421 YSNTEFISDNTSWSAGLHNRNTENSIRGVILDIHFLSQADFLVCTFSQVCRVAYEIMQ 480
DB 421 YPSVEFISDNTSWSAGLHNRNTENSIRGVILDIHFLSQADFLVCTFSQVCRVAYEIMQ 480
QY 481 TLHPDASANFSLDDIYFGGQNAHQIAVPHQRTKEIPMEPGDIIGVAGNHNWYS 540
DB 481 ALHPDASANFSLDDIYFGGPNAHQIAVPHQRTKEIPMEPGDIIGVAGNHNWGY 540
QY 541 KGVNKLKGTGLYPSKYREKTIETVKYTYPEAK 575
DB 541 KGVNKLKGTGLYPSKYREKTIETVKYTYPEADK 575

RESULT 5
ALM1_SCHPO
ID ALM1_SCHPO STANDARD; PRT; 1727 AA.
AC Q9UTK5; O13313; Q9UTR8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Abnormal long morphology protein 1 (Sp8).
GN ALM1 OR SPAC1486.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_Taxid=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Church C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
```

RESULT 6	COG5 DROME	STANDARD;	PRT;	751 AA.
ID	CG05 DROME			
AC	Q9VJD3; Q8TOG3;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Putative conserved oligomeric Golgi complex subunit 5 (Four way stop protein).			
DE	FWS OR CG6549.			
GN	Drosophila melanogaster (Fruit fly).			
OS	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
OX	[1]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=Berkley;			
RC				
RX	MEDLINE=20196006; Pubmed=10731132;			
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Ananatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,			
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,			
RA	Abiril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,			
RA	Balieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,			
RA	Beeson K.Y., Bessu P.V., Bernan B.P., Bhandari D., Bolshakov S.,			
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,			
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,			
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,			
RA	Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,			
RA	Foeller C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,			
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,			
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,			
RA	RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,			
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,			
RA	Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,			
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,			
RA	Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,			
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,			
RA	Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,			
RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,			

RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.; "The genome sequence of Drosophila melanogaster." Science 287:2185-2195(2000).
RL	[2]
RN	SEQUENCE OF 406-751 FROM N.A.
RP	STRAIN=Berkeley; TISSUE=Embryo;
RC	MEDLINE=22426066; PubMed=12537569;
RX	Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M., George R.A., Guarini H., Kronmiller B., Pacleb J.M., Park S., Wan K.H., Rubin G.M., Celnik S.E.; "A Drosophila full-length cDNA resource"; Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
RT	-!- FUNCTION: Required for normal Golgi function (By similarity).
CC	-!- SUBUNIT: Component of the conserved oligomeric Golgi complex which is composed of eight different subunits and is required for normal Golgi morphology and localization (By similarity).
CC	-!- SUBCELLULAR LOCATION: Golgi (By similarity).
CC	-!- SIMILARITY: BELONGS TO THE COG5 FAMILY.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch)
CC	-----
DR	EMBL; AB003655; AAF53619.1; ALT INIT.
DR	EMBL; AY069335; AAL39480.1; FlyBase; Fgn0024689; fws.
KW	Hypothetical protein; Transport; Protein transport; Golgi stack; Membrane.
KW	SEQUENCE 751 AA; 84953 MW; E32B1D05BEEF6EBA CRC64;
SQ	SEQUENCE 751 AA; 84953 MW; E32B1D05BEEF6EBA CRC64;
	Query Match 4.2%; Score 130.5; DB 1; Length 751;
	Best Local Similarity 20.7%; Pred. No. 0.33;
	Matches 127; Conservative 98; Mismatches 212; Indels 177; Gaps 32;
QY	32 DNDHPHSS-----RELSTKLAKLERLKQONDLRRMWSLRPEGP-IDQGATGR 82
DB	19 DNDFTASMHLLTGQIQIBLSK-----QLQNTKEEL---QQVRDKHGALLQQATHAGR 69
QY	83 ----VAVLEEQLVKAKE-----QTENYKKQARND---LGKDHEILRRLRIENGAKELWF 128
DB	70 FDAALNALAEADVQRVRETGHRLKNQVDYQYQVENQTVGLRLHDVSHLLRSAGT---- 124
QY	129 FLQSELKLUKLEGNELORHADEILDLDGH-HERSIMTDLYLYSQDTGAGEWEKEAKOL 187
DB	125 -LLSLTAFLKATK--DVLRILA-EIHPLEGQLIEDKELKDIDFIQQRVAVVISSAQIRNL 180
QY	188 TELVQRRIYYLO--NPKDCSKARKLVCNIN-----KCGYGCQL 224
DB	181 TQM--QLVTGLQERNENVVNALKIPFNFTLEKSIDNLNLTATIAMEOSLKECFAGNDI 238
QY	225 H-----HWVYCFMIAYGTQRT-LILESONRYATGGWET----VFRPVSETCTDRSGI 272
DB	239 SVLNKSPTNVSPAPSRPGKGTPQLTTQNFR--AKFWKSLHWLYDELFEICTQIKLL 296
QY	273 STG-----HMSGEYKDNV-----QVVPELPVIDSLHRPPPLPVLAVPEDLADRLL--- 317
DB	297 KTALEQINGFGYTSESSDQCIPQRFWQQVQLLRKSFDECQPQHVITQLQBELSKLLTSAR 356
QY	318 ----RVHGDPAAVWVSQFKYLIRPQWLBERIEETTK-KLGFPHPIGVGVHVRTDKV-- 370
DB	357 GLQRLHGE-----FQFNEULFAP-----LEVGYYSKCAANFACIAGVDLPNGNEIVDN 405

```

QY 371 -----GTEAHPPIEYMHV-----BEHFOLLERRMKV--DKRVVLTATDDPSLL 414
DB 406 FIRVASTELSAALDSRLTNAIANVPACGKELCTKLEAQIKLGADSKQV---VDLPNIQ 462
QY 415 KEATKYSNYBFISDNSISWSAGLHNRVTENSLRGVLDIHF-----LS 458
DB 463 QQONTQLANVLF-----YKDSVRRMLSDLHVHFEKTPGTAREIISRSLE 507
QY 459 QADFLVCTFSQVCRVAYE---IMQTLHPDASANFHSLD-----DIYFGGQNAH 505
DB 508 QADLLIGTILQIMESIITTSIIVLSMHRPGLNSRMSTGTPSMYNKELQEFVNRWS 567
QY 506 NOIAYVPHQPTKE 519
DB 568 HHIALFDCKQMTKK 581

RESULT 7
MYH9 CHICK
ID MYH9 CHICK STANDARD; PRT; 1959 AA.
AC P1410S;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE MYosin heavy chain, nonmuscle (Cellular myosin heavy chain) (NMWMC).
GN MYH9.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=intestinal epithelium;
RX MEDLINE=90046668; PubMed=2813355;
RA Shohet R.V., Conti M.A., Kawamoto S., Preston Y.A., Brill D.A.,
RA Adelstein R.S.;
RT "Cloning of the cDNA encoding the myosin heavy chain of a vertebrate
cellular myosin.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:7726-7730(1989).
CC -!- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
CAPPING.
CC -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M26510; AAA48974.1; -.
DR PIR; A33977; A33977.
DR HSP; P10587; 1BR2.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; Myosin head.
DR InterPro; IPR004009; Myosin N.
DR InterPro; IPR002928; Myosin_tail.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; Myosin head; 1.
DR Pfam; PF02736; Myosin N; 1.
DR Pfam; PF01576; Myosin tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.

```

```

DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSc; 1.
DR PROSITE; PS00096; IQ; 1.
KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
KW Coiled coil; Alkylation; Multigene family.
FT DOMAIN 1 778 MYOSIN HEAD-LIKE.
FT DOMAIN 779 808 IQ.
FT NP_BIND 837 1925 COILED COIL (POTENTIAL).
FT NP_BIND 174 181 ATP (POTENTIAL).
FT DOMAIN 654 676 ACTIN-BINDING.
FT MOD_RES 694 694 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).
SQ SEQUENCE 1959 AA; 226502 MW; A75C86086FD3A1A1 CRC64;

Query Match 4.1%; Score 127.5; DB 1; Length 1959;
Best Local Similarity 19.8%; Pred. No. 1.8; Indels 189; Gaps 22;
Matches 110; Conservative

QY 42 ELSKILAKLE-RLKQONEDLRMAESLRIPGPIQGTATGRVRLVEELVKAKEQIE-- 98
DB 1066 ELQAIKLELQIQLSKKEELQ--AALARVEEAAQKNAWKKIRELSQITELQEDLESE 1123
QY 99 ----NYKQARNLDGKDEHILRRRIEN-----GAKELWFFLQSELKKLKGNELORH 148
DB 1124 RASRNKAQKQKRDUGLEALEKTELEDTLDSTAQQELRSKREQEVTVLKTLEDEAKTH 1183
QY 149 ADETLDDGHHRSIMTDLVYLSQTDGAGWEKEKDLTELVCORRITYLQNPKDCSKAR 208
DB 1184 EAQI-----QEMRQKHSQAIBELAEQ-----LEQTK 1209
QY 209 KLVCNINKGCGYQQLHHVVVCFMAYGTORTLILESQNWRYATGCGWETVPRPVSETCTD 268
DB 1210 RVKANLEKA-----KQALESER-----AELSNE 1232
QY 269 RSGLTGHWGSEVXDKNV--QVVELPVDLSLHPPVPLVPLAVPEDLADRL--LRVHGDDPA 324
DB 1233 VKVLQKGDAEHKRKKVDAQLOELQVKTGEGER-----VKTLEAERVNKLQVELDNV 1285
QY 325 VWWYSQ----FVKYLIRPQWLEREI-----EETTKLGFKHPVGVHVRDKVGT 372
DB 1286 TGLNLSDSKSIK-LAKDFALESQLOQTOLLEQETRLKLSFS-----TKLKQT 1334
QY 373 EAAPHTPEYVHVVEHFOLLERMMKVKRKYLATDDPSLLKEAKTKYSNYEFTSDNSI 432
DB 1335 EDEKNAKEQLESEEAERKNEKQISVLQQAQV-----EARKQMD-----1374
QY 433 SWSAGLHNRVTENSLRGVLDIHFLSQADFLVCTFSQVCRVAYEIMQTLHPDASANFHS 492
DB 1375 -----DGLGC-----LEIAEAKKKLQKD-----LES 1396
QY 493 LDDIYFGGQNAHNOIAVPHQPTKEIPMEPGDIIIGVAGNHNGYSGVNRKLGKTGL 552
DB 1397 LTQRY-----EKIAAYDKLEKTKRLQOQLDD--IAVDLDHORQTVSNLEKKQKK--- 1445
QY 553 YPSYKVRKEITVXY 567
DB 1446 PDQLAEKNIISAXY 1460

RESULT 8
SCPI RAT
ID SCPI RAT STANDARD; PRT; 997 AA.
AC Q03410;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Synaptonemal complex protein 1 (SCP-1 protein).
GN SYCP1 OR SCPI.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;

```

```

RN  SEQUENCE FROM N.A.
RP  TISSUE=Testis;
RX  MEDLINE=93099884; PubMed=1464329;
RA  Meuwissen R.L.J., Offenber H.H., Dietrich A.J., Rieswijk A.,
RA  Iersel M., Heyting C.;
RT  "A coiled-coil related protein specific for synapsed regions of
RT  meiotic prophase chromosomes.";
RL  EMBO J. 11:5091-5100(1992).
CC  -!- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF
CC  SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS
CC  CHROMOSOMES DURING MEIOTIC PROPHASE.
CC  -!- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF
CC  SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.
CC  FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS
CC  IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE
CC  C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE
CC  SYNAPTONEMAL COMPLEX.
CC  -!- TISSUE SPECIFICITY: Testis.
CC  -!- DEVELOPMENTAL STAGE: EXPRESSED EXCLUSIVELY IN MEIOTIC PROPHASE
CC  CELLS.
CC  -!- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,
CC  FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL
CC  DOMAIN HAS DNA-BINDING CAPACITY.
CC  -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS
CC  CORRECTED IN POSITION 6 TO MAXIMIZE THE SIMILARITY WITH THE
CC  OTHER SPECIES SYCP1 SEQUENCES.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; X67805; CAA48006.1; ALT FRAME.
KW  Nuclear protein; Meiosis; Cell division; Phosphorylation;
KW  DNA-binding; Coiled coil.
FT  DOMAIN 108 819 COILED COIL (POTENTIAL).
FT  DOMAIN 118 121 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT  DOMAIN 701 704 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT  DOMAIN 902 905 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT  DOMAIN 982 990 ARG/LYS-RICH (BASIC).
FT  SEQUENCE 997 AA; 116511 MW; 229D59823FD684BE CRC64;
CC  -----
Query Match 4.1%; Score 127; DB 1; Length 997;
Best Local Similarity 20.3%; Pred. No. 0.8;
Matches 105; Conservative 86; Mismatches 188; Indels 138; Gaps 23;
QY 42 ELSKILAKLERLQKQNEDELRRMAESLRPEGPDQGTATGR-VRVLEEQLVKAQIENY 100
DB 155 ENEKSVLKLEBIEIQENKDL-----IKNNATHWCNLLKTCARSABKTSKY 201
QY 101 K-----KQARNDLGND-----HEILRRRIENGAKELWFFLQSELKKLKGELNQLR 147
DB 202 EVEREBTRQVYVDLNNIEKMILAFELRVQAEARLEMHFKLXEDHEKIQHLE-EYQK 260
QY 148 HAD-----EILLDIGHERSIMTDLYL--SQTDGGEWRKEAKDLTELQVRITVYLQ 199
DB 261 EYNNKENQVSLLLIOSTEKENKMKDLTFLLSESRDKANOLEK-----TKLDENLKEIN 315
QY 200 NPKDCSKARKLVNCKNGKGGYCGQLHHVVCYFMIAVGTORTLILESQNRWYATGGWTFV 259
DB 316 EKXDLHLS-----ELDIKMSQWRSMTOKTL-----EEDLIQIAT---KTIY 354
QY 260 RPVSE-----TCTDRSGLSTGHW----- 277
DB 355 QLTEEKEAQMEELNKAKTTHSLVVTLEKATCTCLEELLRTQORLENNEDQLKLITMELQ 414
QY 278 --SGEVKD---KNQVQVELPIVDSLHPRPPYPLAVPEDLADRLRLRVHG-DPAVWVWSQ 330
DB 415 KKSSELEEMTKFKNNEVELEBKTI-----LAEDQKLLDEKKQVEKLAEBELQKREQ 466

```

```

QY 331 FVKYLI--RPOFWLEREIEETTKLGFHGVHVRRTDKVGTAAAFHPIEYVWVHVE 388
DB 467 ELTFLQLQREKIEHDLLEVQVTVTKTSEEH-----YLKQVEEMKTELEKEKLNIELTANS 521
QY 389 HFOLLERRMKYDKKRVYLATDDPSLLKEAKTKYSNYVEFISDINSWSAGLHNRYTENSIR 448
DB 522 DMLLE-----NKKLVQASDMVLKELKHQEDINCKKOEERMLKQIETLEBK--EMNLR 574
QY 449 GVILDI--HFLSQADFLVCTF--SSQVCR-VAYEIMQ 480
DB 575 DELESVRKEFTQQGDEVKCKLDKSEENARSIEYVLK 611

RESULT 9
SCPI_MESAU STANDARD; PRT; 845 AA.
ID SCPI_MESAU
AC Q60563;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Synaptonemal complex protein 1 (SCP-1 protein) (Meiotic chromosome
DE synapctic protein) (Fragment).
GN SCPI OR SIN1.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=95181577; PubMed=7876343;
RA Dobeon M.J., Pearlman R.E., Karaiskakis A., Spyropoulos B.,
RA Moens P.B.;
RT "Synaptonemal complex proteins: occurrence, epitope mapping and
RT chromosome disjunction.";
RL J. Cell Sci. 107:2749-2760(1994).
CC -!- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF
CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS
CC CHROMOSOMES DURING MEIOTIC PROPHASE (BY SIMILARITY). HAS NON-
CC SPECIFIC DNA BINDING CAPABILITY.
CC -!- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF
CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.
CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS
CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE
CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE
CC SYNAPTONEMAL COMPLEX.
CC -!- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,
CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL
CC DOMAIN HAS DNA-BINDING CAPACITY (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L32978; AAC42039.1; -.
KW Nuclear protein; Meiosis; Cell division; Phosphorylation;
KW DNA-binding; Coiled coil.
FT NON TER 1 672
FT DOMAIN <1 556 COILED COIL (POTENTIAL).
FT DOMAIN 553 556 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 753 756 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 830 838 ARG/LYS-RICH (BASIC).
FT SEQUENCE 845 AA; 99401 MW; D7F28873C824C6A8 CRC64;
SQ
Query Match 4.0%; Score 124.5; DB 1; Length 845;
Best Local Similarity 18.8%; Pred. No. 0.95;

```











```
QY 73 PIDQGTATGRVRLBEEQVKAKEQI-----ENYKQARNDLGKDH----- 112
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 294 KVRKNKLTRELAVLKDELSPAQELNRIEAEKEKFEKEKEKELEHRLKKLQEIKEILK 353
QY 113 -----EILRRINGAKELMFFLOSELKLLKLGNN----- 143
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 354 ELSQLSSSLUKEREYEQAQFEDLSERVEKG-KKLVAETEEKLEKELFSEBEYTSL 412
QY 144 -----ELQGHADLIDLGHHSIMTDLYLVSQTDGAGWEKEAKDLTELVRRI 195
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 413 KKKERLLVQLKELKEKEGQLEN--LTQKI-----KEKKVHEKVLNEL 457
QY 196 TYLQNPDKCSKARKVLCNINKGCGYQLH-HVVYCFM-----IAYGTQRTLILESON 247
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 458 KELERE--LKEREL-----HYHAHWASVLSPGDTCVCGGIYRGKALEN-- 500
QY 248 WRYATGGWETVRPVSCTDRSGSLSTGHWSGEVDKQNVVVELPI-----V 294
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 501 -----VDAEGISELKHAKELKEKEBEREIDTTLKLYAQKINSLEKEEM 541
QY 295 DSIHPRPPVPLAVPDLADRL-----LRVHGDPVAVWVSQFVKYLIRPOPWLER----- 344
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 542 EKLRNEVEELRKEIPENLKERIKLEULRIEKELEHKLKLYRKALEDRQKQKEAQAQL 601
QY 345 -----EIEETTKKLGKFPVIG-----VHVRRTDKVG-----TEAAPHPIEYVMVH 385
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 602 HKAQTELELLKEKIREKSLVKFEKLYRVERLEDYEESLKEEINVINSKLQEIKEKEKK 661
QY 386 VEEHF-OLLERMVVDKRVLYATDDPSLLKKAQTK 420
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 662 LRKHFEELSSRSKLEGEALSALNESINSLEERKEK 697
```

```
RESULT 13
ITN1_XENLA
ID ITN1_XENLA STANDARD; PRT; 1270 AA.
AC O42287;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Intersectin 1.
GN ITN1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Oocyte;
RX MEDLINE=99030416; PubMed=9813051;
RA Yamabhai M., Hoffman N.G., Hardison N.L., McPherson P.S.,
RT Castagnoli L., Cesareni G., Kay B.K.;
RT "Intersectin, a novel adaptor protein with two eps15 homology and five
RT src homology 3 domains";
RL J. Biol. Chem. 273:31401-31407(1998).
CC -!- FUNCTION: Adaptor protein that may provide indirect link between
CC the endocytic membrane traffic and the actin assembly machinery.
CC May regulate the formation of clathrin-coated vesicles.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; membrane-associated protein.
CC Enriched in synaptosomes (By similarity).
CC -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC -!- SIMILARITY: Contains 5 SH3 domains.
CC -!- SIMILARITY: Contains 2 EH domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL; AF032118; AAC73068.1; -
DR PIR; T09194; T09194.
DR HSSP; P29355; 1SEM.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000261; EPS15 homology.
DR InterPro; IPR000108; Neu_cyt_fact_2.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00036; ehand; 2.
DR PRINTS; PR00018; SH3; 5.
DR PRINTS; PR00499; P67PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRODOM; PD000066; SH3; 5.
DR SMART; SM00054; EFh; 2.
DR SMART; SM00027; EH; 2.
DR SMART; SM00326; SH3; 5.
DR PROSITE; PS00018; EF HAND; 2.
DR PROSITE; PS50031; EH; 2.
DR PROSITE; PS50002; SH3; 5.
DR Endocytosis; SH3 domain; Repeat; Coiled coil; Calcium-binding.
KW DOMAIN 21 109
FT CA BIND 66 77 EF-HAND 1 (POTENTIAL).
FT DOMAIN 220 309 EH 2.
FT CA BIND 266 277 EF-HAND 2 (POTENTIAL).
FT DOMAIN 325 697 LYS/LEU/GLU/ARG/GLN-RICH (KLERQ).
FT DOMAIN 349 691 COILED COIL (POTENTIAL).
FT DOMAIN 732 793 SH3 1.
FT DOMAIN 897 955 SH3 2.
FT DOMAIN 986 1044 SH3 3.
FT DOMAIN 1058 1122 SH3 4.
FT DOMAIN 1139 1198 SH3 5.
SQ SEQUENCE 1270 AA; 143670 MW; EA940C1F6B6A6858 CRC64;

Query Match 3.9%; Score 119.5; DB 1; Length 1270;
Best Local Similarity 20.6%; Pred No. 3.5;
Matches 121; Conservative 82; Mismatches 174; Indels 211; Gaps 33;

QY 41 RELSKILAKLERLKQNNEDLRMAESLRIPGIDQGTATGRVRLBEEQVKAKEQIENY 100
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 375 RKEQERLAQLERAEOEKERERQOERK-----RQDLEKQLEKQRE-LERQ 420
QY 101 KQARNDLGKDHILRRRIENGAKELMFFLOSELKLLKLGNNELQGHADLIDLGHHE 160
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 421 REBERR-----KEIERR--EAKR-----ELERQLEWERNR--QELL 456
QY 161 RSIWTDLYLSQTDGAGWEKEAKDLTELVRRI-----LQNPKDC--SKARKLVC 212
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 457 -----NQRNEQEDIVLKAKKKTLEFELEALNDKHKHLEGLQDIRC 499
QY 213 NINRGCGYCOLHHVVCYFMIAVGTQRTLILESONRYATCGWETVFRPVSETCTDRSGL 272
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 500 RL-----TORHEI-ESTN----- 512
QY 273 STGHWSGEVDKQNVVVELP-IVDSLHPRPPVPLAVPE--DLADRLRV-----HGDP 324
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 513 -----KSRELRIAEITHLQOQLQESQOLLGKMIPEKOSLIDOLKQVQONSILHRDSL 563
QY 325 VWWSQFVKYLIRPOPWLERIEETTKKLGKFPVIGVHVR-RTDKV--GTEAAPHPIEE 381
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 564 L-----TLKRALE--TKEIRGQ-----QLRDQDLEVEKETRAKLEIDV 600
QY 382 YMVHVEHFQLLERRM-----KVDKRVVYLATDDPSLLKKAQTKSYNYEFSINS 431
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 601 FNNQLKELRYLNKQFQKQDFETEKIKQELERKTSSELDKLEEDKR-----RMLEQDK 656
QY 432 ISWAGLHNRYTNSLRGVLDIHLFSLQADFLVCTFSQVCRVAYEIMQTLHPDASAFH 491
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 657 L-W-----QDRVKQBEERYKFDQBEKEKEE-----SVQKCEV--EKKEPQEKPNKPFH 703
QY 492 SLDDIIVYFGGQ-----NAH-----NQ-----IAVPHQPRTKKEIPMPGDIIGVAG 533
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 704 QPPRPGKGGQIPWNTTEKAPLTINQGDVKKVYVRYALYPPDARSHDEITIEFGDILWDE 763
```

QY 534 NH-----WNGYSGVNRKLGKGTGLYPSYKVRKIEYKYP--TYPEAE 574  
 Db 764 SQTGPGWLG-----GELKGTGWFFA-NYAERMPSEFPSTTKPAE 805

## RESULT 14

SCPI\_HUMAN  
 ID SCPI\_HUMAN STANDARD; PRT; 976 AA.  
 AC Q15431; O14963;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 16-OCT-2001 (Rel. 36, Last sequence update)  
 DE Synaptonemal complex protein 1 (SCP-1 protein).  
 GN SYCP1 OR SCPI.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=97224467; PubMed=9119375;  
 RA Meuwissen R.L.J., Meerts I., Hoovers J.M.N., Leschot N.J.,  
 RA Heyting C.;  
 RA "Human synaptonemal complex protein 1 (SCP1): isolation and  
 RT characterization of the cDNA and chromosomal localization of the  
 RT gene";  
 RL Genomics 39:377-384 (1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=98037449; PubMed=9371398;  
 RA Kondoh N., Nishina Y., Tsuchida J., Koga M., Tanaka H., Uchida K.,  
 RA Inazawa J., Taketo M., Nozaki M., Nojima H., Matsumiya K., Namiki M.,  
 RA Okuyama A., Nishimune Y.;  
 RT "Assignment of synaptonemal complex protein 1 (SCPI) to human  
 RT chromosome 1p13 by fluorescence in situ hybridization and its  
 RT expression in the testis";  
 RL Cytogenet. Cell Genet. 78:103-104 (1997).  
 CC -!- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF  
 CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS  
 CC CHROMOSOMES DURING MEIOTIC PROPHASE.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF  
 CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.  
 CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS  
 CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE  
 CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE  
 CC SYNAPTONEMAL COMPLEX (BY SIMILARITY).  
 CC -!- TISSUE SPECIFICITY: Testis.  
 CC -!- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,  
 CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL  
 CC DOMAIN HAS DNA-BINDING CAPACITY (BY SIMILARITY).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcements>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X95654; CAA64956.1; -;  
 CC EMBL; D67035; BAA22586.1; -;  
 CC GenBank; HGNC:11487; SYCP1.  
 CC MIM; 602162; -;  
 CC GO; GO:0005634; C:nucleus; TAS.  
 CC GO; GO:0005716; C:synaptonemal complex; TAS.  
 CC GO; GO:0003677; F:DNA binding activity; TAS.  
 CC GO; GO:0007131; P:meiotic recombination; TAS.  
 CC GO; GO:0007283; P:spermatogenesis; TAS.  
 CC GO; GO:0007129; P:synapsis; TAS.  
 CC Nuclear protein; Meiosis; Cell division; Phosphorylation;  
 CC DNA-binding; Coiled coil.

FT DOMAIN 12 100  
 FT DOMAIN 107 798  
 FT DOMAIN 117 120  
 FT DOMAIN 679 682  
 FT DOMAIN 880 883  
 FT DOMAIN 961 969  
 FT CONFLICT 46 46  
 FT CONFLICT 106 106  
 FT CONFLICT 153 153  
 FT CONFLICT 161 161  
 FT CONFLICT 168 168  
 FT CONFLICT 216 216  
 FT CONFLICT 225 226  
 FT CONFLICT 350 350  
 FT CONFLICT 360 360  
 FT CONFLICT 400 401  
 FT CONFLICT 406 406  
 FT CONFLICT 415 415  
 FT CONFLICT 449 449  
 FT CONFLICT 483 510  
 FT CONFLICT 516 528  
 FT CONFLICT 549 549  
 FT CONFLICT 560 560  
 FT CONFLICT 805 805  
 FT CONFLICT 941 941  
 SQ SEQUENCE 976 AA; 111069 MW; 8BA81D042AC2696B CRC64;  
 Query Match 3.8%; Score 118.5; DB 1; Length 976;  
 Best Local Similarity 20.5%; Pred. No. 2.9;  
 Matches 90; Conservative 74; Mismatches 164; Indels 111; Gaps 19;  
 QY 42 ELSKILAKLEBLKQONEDLRMAESLRIPEDIGDTATGR---VRVLEEQLVKAQEIQE 98  
 Db 435 ELKVLGKETLLYENKQFEKIAELKGTQEQL-IGLQAREKEVHDLFIQLTJTTSQ 493  
 QY 99 NYKQARNDLGKDHRIIRRIENGAKELWFFLOSELKKLKLEGNELQORHDAEILLDLGH 158  
 Db 494 YSKEVKD-----LATELEN-EKLNKLTSHCNKL-SLENKELTQETSDMTLELKN 543  
 QY 159 HERSI-----MTDLYLSQTD-----GAGENWEKEAKDLTELVRITRYLQNPDC 204  
 Db 544 QQEDINNNKQEBMLKQIENLQETETQLRNELEYVREELKQKRDVCKLD--KSEENC 601  
 QY 205 SKARKLVGNIN-----KCGYGCQLHHVYCFMIAYGTQRTILLESQNRW 249  
 Db 602 NNLKQVENKNKYIEELOQENKALKKKGTAESKQLN--VYEIKV---NKLELELESQK 656  
 QY 250 YATGQWTVPRVSETCTDRSLSTGHWSGVKDKNQVQVVELPIVDLSLHPRPYLPLAVP 309  
 Db 657 F-----GEITDTYQKIEDKKIS-----E 675  
 QY 310 EDLADRL--LRVHSDPAV-----WVVSQFVKYLIRPQWLERIEETTKLGF-F 355  
 Db 676 ENLLEVEKAKVIADAEAVLKQEKDKRCQKIAEMVALMEKHQYDKIIEERDSELGLY 735  
 QY 356 KHPVIGVHVRVDKVGTEAAAFHPHIEYVMVHEHPOLLERMKVDKRYLVTDDPSLLK 415  
 Db 736 KSK---EQEQSLASLELSNLKAEILLSVKQKLE-IEREEKEKLKRE--AKENTATLK 789  
 QY 416 EAKTKSYNVEFISDINSISW 434  
 Db 790 EKKOKKTTOTFLLEPFIYW 808  
 RESULT 15  
 ID REST\_HUMAN STANDARD; PRT; 1427 AA.  
 AC P30622;  
 DT 01-APR-1993 (Rel. 25, (reated)  
 DT 01-APR-1993 (Rel. 25, last sequence update)  
 DT 28-FEB-2003 (Rel. 41, last annotation update)

DE Restin (Cytoplasmic linker protein-170 alpha-2) (CLIP-170) (Reed-  
DE Sternberg intermediate filament associated protein).  
GN RSN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Peripheral blood monocytes;  
RX MEDLINE=92289675; PubMed=1600942;  
RA Bilbe G., Delabie J., Bruegggen J., Richener H., Asselbergs F.A.M.,  
RA Cerletti N., Sorg C., Odink K., Tarcsay L., Wieseandanger W.,  
RA de Wolf-Peeters C., Shipman R.;  
RT "Restin: a novel intermediate filament-associated protein highly  
RT expressed in the Reed-Sternberg cells of Hodgkin's disease,"  
RL EMBO J. 11:2103-2113(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92405160; PubMed=1356075;  
RA Pierre P., Scheel J., Rickard J.E., Kreis T.E.;  
RT "CLIP-170 links endocytic vesicles to microtubules.";  
RL Cell 70:887-900(1992).  
CC -!- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN  
CC THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE  
CC CYTOSKELETON.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=alternative splicing; Named isoforms=2;  
CC Name=Long;  
CC IsoId=P30622-1; Sequence=Displayed;  
CC Name=Short;  
CC IsoId=P30622-2; Sequence=VSP 000765;  
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBERG CELLS  
CC OF HODGKIN'S DISEASE.  
CC -!- SIMILARITY: Contains 2 CAP-Gly domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X64838; CAA46050.1; -;  
DR EMBL; M97501; AAA35693.1; -;  
DR PIR; S22695; S22695.  
DR Genbank; HGNC:10461; RSN.  
DR MIM; 179838; -;  
DR GO; GO:0005768; C:endosome; TAS.  
DR GO; GO:0005882; C:intermediate filament; TAS.  
DR GO; GO:0015630; C:microtubule cytoskeleton; TAS.  
DR GO; GO:0008017; F:microtubule binding activity; TAS.  
DR GO; GO:0006899; P:non-selective vesicle transport; TAS.  
DR InterPro; IPR000938; CAP-Gly.  
DR InterPro; IPR001878; Znf.CCHC.  
DR Pfam; PF01302; CAP\_GLY\_2.  
DR SMART; SM00343; Znf.C2HC; 1.  
DR PROSITE; PS00845; CAP\_GLY\_1; 2.  
DR PROSITE; PS0245; CAP\_GLY\_2; 2.  
DR Cytoskeleton; Microtubules; Coiled coil; Repeat; Alternative splicing.  
FT DOMAIN 78 120 CAP-GLY 1.  
FT DOMAIN 143 204 SER-RICH  
FT DOMAIN 232 274 CAP-GLY 2.  
FT DOMAIN 304 331 SER-RICH.  
FT DOMAIN 350 1342 COILED COIL (POTENTIAL).  
FT DOMAIN 1408 1421 CCHC-BOX.  
FT VARSPLIC 457 491 Missing (in isoform Short).  
FT FTID=VSP 000765.  
FT CONFLICT 1069 1069 D -> E (IN REF.2).  
FT SEQUENCE 1427 AA; 160989 MW; 0A4F16DD94254E8 CRC64;

Query Match 3.8%; Score 118.5; DB 1; Length 1427;  
Best Local Similarity 17.9%; Pred. No. 4.7; Indels 219; Gaps 26;  
Matches 117; Conservative 104; Mismatches 215;  
QY 40 SRELSKILAKLERLKQONEDL-----RRMAESLRIPEGPIDQGTATGRVR 84  
DB 598 SKENESKSLKLEHANKENSIVIALWKSLETAIAHQQAAMEELKVSPSK-GLGTETAFA 646  
QY 85 VLEBQLVKAK-----EQIENYKQARNDLGKHDEILRRRIENGAKELWFFLOSELK 137  
DB 647 ELKTQIEKWRLDYQHEIENLQNDQDSERAAHAKAMEALRAKLMKVKEKENSLEAIRSK 706  
QY 138 KKLLEG-----NELQRHADEILLDLGHHSIMTDL 167  
DB 707 DKAEDQHLVEMEDTLNKLQEAIEIKVKELEVLQAKCNEQTKVIDNFTSOLKATEEKL-DL 765  
QY 168 YLQSOTDAGWEKREKADLTVELVORRITYLQNPK--DCSKARKL-----VCNIN 215  
DB 766 DALRKASSEKSEMKKLRQQLAAEKQIKHLEIKNAESSKASITRELOQRELKLTNLQ 825  
QY 216 KCGYGCQLHHVYCFMIAYQRTLILESQNMRYATGWTETPRPVSSETCTDRSGLSTG 275  
DB 826 ENLSEVSQVKETL-----EKELQILKEKFAEASEAVSVQSMQET----- 866  
QY 276 HWSGEVDKQNVQVELPIVDSLHPPPYLPL-----AVPEDLAD----- 314  
DB 867 -----VNLHQKEEQFNMLSSDLEKLRNLADMEAKFREKDEREQ 908  
QY 315 -----RLLRVHGDPVWVWQFVKYLIRPOPWL-EREITEETTKLGFKHPVIG 361  
DB 909 IKAKEKLENDIAEIMKMGDNS-----SOLTK--MNDELRLKERDVEELQLKL----- 954  
QY 362 VHVVRTDKVGTAAAF--HPIEEYMHVVEHFPQLLERRMKVDKRVYLATDPPSLLEAKT 419  
DB 955 -----TKANENASFLQKSIEDMTVKAESQQAQKHEEKELERKLSLSD--LEKKMET 1006  
QY 420 KYSNYEFIS-----DNSISWSAGLHRY-----TENSRL-----GVILDIHFL-S 458  
DB 1007 SHNQCOELKARYERATSETTKHBEILQNLQKTLTDEKLGKARENSGLLEELRK 1066  
QY 459 QADFLVCTFSQVCRVAEIMQTLHPDASANFSLDDIYFEGG--QN-----AHNQIAVY 511  
DB 1067 QAD---KAKAAQTADADAMQIWEQTKETETLASLETQTNAKLQNELDTLKNNKXV 1123  
QY 512 PHQPRTKEEIPMEPGDIIGVAGNHWGYSKGVNRKLGKTKGLYPSYKVKREKTIETVK 566  
DB 1124 EELNKSKELTVE-----NQKM-----EPRKEIETLK 1151

Search completed: February 2, 2004, 08:40:18  
Job time : 14.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 08:39:43 ; Search time 33.5 Seconds  
(without alignments)  
4429.261 Million cell updates/sec

Title: US-09-971-773-23  
Perfect score: 3081  
Sequence: 1 MRATGSGRWIMLILFANGT.....YKREKIETVKYPTYPEAK 575

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:\*

- 1: sp\_arChaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2783.5	90.3	578	13 Q8AXS7	Q8axs7 xenopus lae
2	2176	70.6	446	4 Q8NEP2	Q8nep2 homo sapien
3	1489.5	48.3	619	5 Q9YVY5	Q9yvy5 d putative
4	991	32.2	559	5 O16882	O16882 caenorhabdi
5	810	26.3	169	4 Q8IUAS	Q8iuas homo sapien
6	671	21.8	512	5 Q8IFW9	Q8ifw9 ciona intes
7	614.5	19.9	560	5 Q8IFW8	Q8ifw8 ciona intes
8	137.5	4.5	1137	16 Q8F3E7	Q8f3e7 leptospira
9	132.5	4.3	876	17 Q8TX14	Q8tx14 methanopyru
10	128	4.2	674	5 Q18105	Q18105 caenorhabdi
11	127.5	4.1	1218	5 Q8IED2	Q8ied2 plasmodium
12	124	4.0	1025	10 Q9SAAS	Q9saas arabidopsis
13	124	4.0	1285	4 Q9UEG2	Q9ueg2 homo sapien
14	124	4.0	2442	4 O14812	O14812 homo sapien
15	122.5	4.0	681	11 Q8BU18	Q8bul18 mus musculu
16	122.5	4.0	1200	11 Q921B9	Q921b9 mus musculu

17	122	4.0	1871	5 Q8IR54	Q8ir54 drosophila
18	122	4.0	2328	5 Q9VY43	Q9vy43 drosophila
19	122	4.0	2360	5 Q8IR55	Q8ir55 drosophila
20	122	4.0	2501	5 Q9NCW7	Q9ncw7 drosophila
21	121.5	3.9	975	13 Q98TQ5	Q98tq5 notothenia
22	121.5	3.9	3187	11 Q63714	Q63714 rattus norv
23	121	3.9	1285	4 Q9BZZ7	Q9bzz7 homo sapien
24	120.5	3.9	584	17 Q96YQ6	Q96yq6 sulfolobus
25	120.5	3.9	3542	5 Q9USM2	Q9usm2 plasmodium
26	119.5	3.9	612	4 Q96JV2	Q96jv2 homo sapien
27	119.5	3.9	2007	13 Q02015	Q02015 gallus gall
28	119.5	3.9	3616	13 Q9W6V0	Q9w6v0 gallus gall
29	119	3.9	1871	10 Q9SRD5	Q9srd5 arabidopsis
30	118	3.8	581	5 Q21065	Q21065 caenorhabdi
31	118	3.8	717	4 Q96ED9	Q96ed9 homo sapien
32	118	3.8	1999	11 Q63731	Q63731 rattus norv
33	117.5	3.8	1265	3 Q59920	Q59920 pneumocysti
34	117.5	3.8	1397	10 Q9LP90	Q9lp90 arabidopsis
35	117.5	3.8	1929	13 Q98TQ6	Q98tq6 notothenia
36	117	3.8	1177	16 Q8RCY8	Q8rcy8 thermoanaer
37	116	3.8	178	4 Q8WX78	Q8wx78 homo sapien
38	116	3.8	214	6 Q8MJ49	Q8mj49 sus scrofa
39	116	3.8	217	4 Q96I04	Q96i04 homo sapien
40	116	3.8	283	10 Q9C717	Q9c717 arabidopsis
41	116	3.8	1156	16 Q66878	Q66878 aquifex aeo
42	116	3.8	1871	5 Q9NCL3	Q9nc13 drosophila
43	116	3.8	2442	4 Q9H450	Q9h450 homo sapien
44	116	3.8	2442	4 Q60588	Q60588 homo sapien
45	115.5	3.7	716	11 Q8BY47	Q8by47 mus musculu

#### ALIGNMENTS

#### RESULT 1

Q8AXS7 PRELIMINARY; PRT; 578 AA.  
AC Q8AXS7;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Putative alpha-6-fucosyltransferase (EC 2.4.1.68).  
GN FUT8.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Mollicone R., Michalski J.C., Bauvy C., Cailleau-Thomas A., Oriol R.;  
RA Candelier J.J., Martinez-Duncker I., Breton C., Codogno P., Oriol R.;  
RT "Splice variants of alpha6-fucosyltransferase are expressed early in human embryogenesis."  
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ514872; CAD55853.1; -  
KW Transferase; Glycosyltransferase.  
SQ SEQUENCE 578 AA; 66578 MW; 2741A03BDB3C5265 CRC64;

Query Match 90.3%; Score 2783.5; DB 13; Length 578;  
Best Local Similarity 87.7%; Pred. No. 2e-192;  
Matches 507; Conservative 40; Mismatches 28; Indels 3; Gaps 1;  
Qy 1 MRATGSGRWIMLILFANGTLLFYIGGHLVRDNDPHSSRELKILAKLERLKOQNDL 60  
Db 1 MRATGSGRWIMLILFANGTLLFYIGGHLVRDNDPHSSRELKILAKLERLKOQNDL 60  
Qy 61 RRMASRIPEGFTDOGTATGRVRLVLEQLVKAKEQLENYKKOARN---DLGKDEILRR 117  
Db 61 RRMASRIPEGFTDOGTATGRVRLVLEQLVKAKEQLENYKKOARN---DLGKDEILRR 117  
Qy 118 RIENGAKELWFFIQSELKLLKLEGNLQRLHDEILLDLGHHRSIMTDLYYLSQTDGAG 177  
Db 118 RIENGAKELWFFIQSELKLLKLEGNLQRLHDEILLDLGHHRSIMTDLYYLSQTDGAG 177

```
Db 121 AIEGAKBFWYLVQSEVKKLHLDRLNQLRHVDEILIDMGHQORSVMTDLYLSQTDGAG 180
Qy 178 EWEKEAKDLTELVRITLYLQNPCKSKARKLVNINKGCGYQGLHHVYVCFMAYGT 237
Db 181 DWREREAKDLTLVQRITLYLQNPCKSKARKLVNINKGCGYQGLHHVYVCFMAYGT 240
Qy 238 QRTLLSQNRWYATGCHETPRPVSETCTDRSGLSTGHWSGEVCKDNVQVPELPIVDSL 297
Db 241 QRTLLSQNRWYATGCHETPRPVSETCTDRSGLSTGHWSGEVCKDNVQVPELPIVDSL 300
Qy 298 HRPYPPLPLAVPEDLADLRLLRHGDPVWVWSQFVKYLIRPQWLERIEETTKLGFPH 357
Db 301 HRPYPPLPLAVPEDLADLRLLRHGDPVWVWSQFVKYLIRPQWLERIEETTKLGFPH 360
Qy 358 PVIGVHVRTDKVGTAAFPHEIEEYVWVHEHFQLLERRMKVDKRVYLATDDPSLLKEA 417
Db 361 PVIGVHVRTDKVGTAAFPHEIEEYVWVHEHFQLLERRMKVDKRVYLATDDPSLLKEA 420
Qy 418 KTKYNYEFISDNSISWSAGLHNRYTENSLRGVILDIHFLSQADFLVCTFSQVCRVAYE 477
Db 421 KAKYQYEFISDNSISWSAGLHNRYTENSLRGVILDIHFLSQANFLVCTFSQVCRVAYE 480
Qy 478 IMQTLHPDASANFSLDDIYFPGQNAHNOIAVYPHQRTKEEIPMEPCDIIIGVAGNHN 537
Db 481 IMQTLHPDASAFHSLDDIYFPGQNAHNOIAVYPHQRTKEEIPMEPCDIIIGVAGNHN 540
Qy 538 GYSKGNRKLGTGLYPSYKVKREKTIETVKYPTYPEAEK 575
Db 541 GYSKGNRKLGTGLYPSYKVKREKTIETVKYPTYPEAEK 578

RESULT 2
Q8NEP2
ID Q8NEP2 PRELIMINARY; PRT; 446 AA.
AC Q8NEP2;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DE Putative alpha-(1,6)-fucosyltransferase (EC 2.4.1.68) (Glycoprotein 6-
DE alpha-L-fucosyltransferase) (GDP-fucose-glycoprotein
DE fucosyltransferase) (GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alpha1,6-
DE fucosyltransferase) (Alpha1-6FucT).
GN FUC6 OR CG2448.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George K.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foaier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matti B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
```

```
Qy 401 KKRVLATDDPSLLKEAKTKYSNYEFISDNSISWSAGLHNRYTENSLRGVILDIHFLSQ 460
Db 272 KKRVLATDDPSLLKEAKTKYPNYEFISDNSISWSAGLHNRYTENSLRGVILDIHFLSQ 331
Qy 461 DFLVCTFSQVCRVAYEIMOTLHPDASANFSLDDIYFPGQNAHNOIAVYPHQRTKEE 520
Db 332 DFLVCTFSQVCRVAYEIMOTLHPDASANFSLDDIYFPGQNAHNOIAVYPHQRTKEE 391
Qy 521 IPMEPCDIIIGVAGNHNWYSGKGNRKLGTGLYPSYKVKREKTIETVKYPTYPEAEK 575
Db 392 IPMEPCDIIIGVAGNHNWYSGKGNRKLGTGLYPSYKVKREKTIETVKYPTYPEAEK 446

RESULT 3
Q9VVV5
ID Q9VVV5 PRELIMINARY; PRT; 619 AA.
AC Q9VVV5;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Putative alpha-(1,6)-fucosyltransferase (EC 2.4.1.68) (Glycoprotein 6-
DE alpha-L-fucosyltransferase) (GDP-fucose-glycoprotein
DE fucosyltransferase) (GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alpha1,6-
DE fucosyltransferase) (Alpha1-6FucT).
GN FUC6 OR CG2448.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George K.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foaier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matti B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
```

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley; TISSUE=Testis;  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nuncio J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP HOMOLOGY.  
 RX PubMed:1698403;  
 RA Roos C., Kolmer M., Mattila P., Renkonen R.;  
 RT "Composition of *Drosophila melanogaster* proteome involved in  
 RT fucosylated glycan metabolism."  
 RL J. Biol. Chem. 277:3168-3175(2002).  
 CC -!- FUNCTION: CATALYZES THE ADDITION OF FUPOSE IN ALPHA 1-6 LINKAGE TO  
 CC THE FIRST GLCNAC RESIDUE, NEXT TO THE PEPTIDE CHAINS IN N-GLYCANS  
 CC (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: GDP-L-FUCOSE + N4-(N-ACETYL-BETA-D-  
 CC GLUCOSAMINYL-1,2-ALPHA-D-MANNOSYL-1,3-(R(1)-ALPHA-1,6)-BETA-D-  
 CC MANNOYL-BETA-N-ACETYL-1,4-D-GLUCOSAMINYL-1,4-N-ACETYL-D-  
 CC GLUCOSAMINYL)ASPARAGINE = GDP + N4-(N-ACETYL-BETA-D-GLUCOSAMINYL-  
 CC 1,2-ALPHA-D-MANNOSYL-1,3-(R(1)-ALPHA-1,6)-BETA-D-MANNOSYL-1,4-  
 CC BETA-N-ACETYL-D-GLUCOSAMINYL-1,4-(ALPHA-L-FUCOSYL-1,6)-N-ACETYL-D-  
 CC GLUCOSAMINYL)ASPARAGINE.  
 CC -!- PATHWAY: GLYCOSYLATION.  
 CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND  
 CC FORM IN TRANS CISTERNAE OF GOLGI (BY SIMILARITY).  
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 23.  
 DR EMBL; AF441264; AAN63649.1; -;  
 DR EMBL; AE003487; AAF48079.1; -;  
 DR EMBL; AY051451; AAK92875.1; -;  
 DR FlyBase; FBgn0030327; CG2448.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF00018; SH3; 1.  
 DR SMART; SM00326; SH3; 1.  
 KW Hypothetical protein; Transferase; Glycosyltransferase; Transmembrane;  
 KW Signal-anchor; Golgi stack; SH3 domain.  
 FT DOMAIN 1 17  
 FT TRANSMEM 18 38  
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT LUMENAL, CATALYTIC (POTENTIAL).  
 FT SH3.  
 FT SH3-BINDING (POTENTIAL).  
 FT IMPORTANT FOR DONOR SUBSTRATE BINDING  
 FT (BY SIMILARITY).  
 SQ SEQUENCE 619 AA; 70205 MW; 691BFD9B5C6557DE CRC64;  
 Query Match 48.3%; Score 1489.5; DB 5; Length 619;  
 Best Local Similarity 46.6%; Pred. No. 5e-99;  
 Matches 285; Conservative 110; Mismatches 160; Indels 49; Gaps 7;  
 QY 3 AWTGSW-RWIMLILFAWGTLLFYIGHLVRDN-----DHPDSSRELKILAKLERKQQ 56  
 DB 10 ASANSWALALIFVLAWGLVTVFVVKLTNTGQQAAGESELNARRISQALOMLEHTQR 69  
 QY 57 NEDLRMAESLRIPGPIDQGTGRVRLVEQLVKAKEIE-----  
 DB 70 NEELKQLIDELMSDQ--LDKQSAKLVQLRLDNLNPKLAPVAGPEPESMFESAPADLR 127  
 QY 99 ---NYKKQARNDL-----GKQHEILRRRIENGAKELWPFQSELKUKKL--- 140  
 DB 128 GWNVAEGAPNDLEAGVPDHGEFEPSELEVFTRRIQTNIGIWMFFSELGKVRKAVAA 187  
 QY 141 --EGNELQRADEIILLDIGHHSRTMDLYLSQTDGAGEWREKAKDLTELQVRITYL 198

188 GHASADLEESINQVLLQCAEHKRSLLSDMERMRQSDGYEARHKEARDLSLVQRRLLHL 247  
 QY 199 QNPDKCSKARKLVNINKGCGYCOLHHVVCYFMIAVGTORTLLESQNRVYATGGWETV 258  
 DB 248 QNPSDCQARKLVCKLNKCGYGCQLHHVVCYFIVAYATERTLLKSRGWRHYHKGWEEV 307  
 QY 259 FRPVSETCTDRSGLSTGHWSGEVKDKNVQVVELPIVDSLHPRPPYPPLVPLAVPEDLADRLR 318  
 DB 308 FQVSNNSCHDAGTANTYNWPK---PNTQVLVLPIDSLMPPYPPLVPLAVPEDLAPLKR 364  
 QY 319 VHGDPAVWVSQFVKYLIIRPQWLERIEBETTKLGPVKHVGHVHRTDKVGTAAAPHP 378  
 DB 365 LHGDPIVWVWQFLKYLRLPQPTTRDFLTSGMRNLGWERPIVGVHVHRTDKVGTAAACHS 424  
 QY 379 IEYVHVVEHFPQLERRMKVDKRVYLATDDPSLLKEAKTKYSNYEFISNSISWSAGL 438  
 DB 425 VEYMTYVEDYRTLEVNGSTVAARIIFLASDDAQVIEARRKYPQYQLIGPEVARMASV 484  
 QY 439 HNYTENSRLGVILDIHFLSOADFLVCTFSQVCRVAYEIMOTLHPDASANFHSLLDDIY 498  
 DB 485 STRYTDALNGILDIHLLSNSDLVCTFSQVCRVAYEIMQWYPDAAHFRKSLDDIY 544  
 QY 499 FGGQNAHQIAVPHQPTKEIPMEPGDIIIGVAGNHNGYSKGNKRLGKTGLYPSYKV 558  
 DB 545 YGGQNAHNRVVAHAKPRTHEDLQRLVGDLSVAGNHWGDSGKGNKTRTNQGLFPSPKV 604  
 QY 559 REKIEVTKYPTY 570  
 DB 605 EEKVDYAKPLY 616

RESULT 4  
 ID 016882 PRELIMINARY; PRT; 559 AA.  
 AC 016882;  
 DT 01-JAN-1998 (TReMBLrel. 05, Created)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
 DE Hypothetical protein C10F3.6 (Core alpha-6-fucosyltransferase)  
 DE (EC 2.4.1.68).  
 GN C10F3.6 OR FUT8.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;  
 OC Rhabditidae; Peloderidae; Caenorhabditis.  
 OK NCBI\_taxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=95069613; PubMed=9851916;  
 RA Waterston R.;  
 RT "Genome sequence of the nematode *C. elegans*: a platform for  
 RT investigating biology. The *C. elegans* Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Davidson S., Wohlmann P.;  
 RT "The sequence of *C. elegans* cosmid C10F3.";  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Balazino L., Oriol R., Michalski J.C., Candelier J.J.,  
 RA Martinez-Duncker I., Mollicone R.;  
 RT "Cloning, expression and genomic organization of two core  
 RT fucosyltransferases (CSE and Ced) from *Caenorhabditis elegans*."  
 RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF022968; AAN84870.1; -;



[illegible]





Qy 399 VDKRRVLTDDPSLLKAKTKYNGYEFISDMSISWSAGLHRYTEN 445  
Db 626 AVEKRTAAKSLDMSLRKTE-----IDYEDIKTESIRLQEEYNNMLLTN 668

## RESULT 11

Q8IED2 PRELIMINARY; PRT; 1218 AA.  
AC Q8IED2; (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Chromosome segregation protein, putative.  
GN MAL13P1.96.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Harris B., Lennard N., Clark L., Line A., Barton A., Corton C.,  
RA Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,  
RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.,  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL844509; CAD52328.1;  
SQ SEQUENCE 1218 AA; 143059 MW; E1DBD7838BBC7E5B CRC64;

Query Match 4.1%; Score 127.5; DB 5; Length 1218;  
Best Local Similarity 24.8%; Pred. No. 2.4;  
Matches 60; Conservative 41; Mismatches 52; Indels 89; Gaps 14;

Qy 41 RELSKILAKLER-----LQONEDLRRMAESLRPEPIDQGTATGRVVRVLEEQIVK 92  
Db 784 KKLTEVIRKLEKIDSEYANKDKKEEDLKETIKLK-----NKKIQLETEBEHK 831  
Qy 93 AKE-----QIENYKQAR---NDLGKDHILR-----RRIENG 122  
Db 832 KKEEIDVLLQIENYKKQKKEETNDLSTDEINEIEKKIEDIEKNINITRENKLELENK 891  
Qy 123 AKEL---WFFQSELEK-LKLEGNELQORHADEILDLGHHSRSTMTDLY----- 168  
Db 892 ITELQSSFSSEYENEMKHVVKKIE--DLEKKSENILDLKULENTLL-DLQKDLKTSSTV 948  
Qy 169 -YLSOTDGCAGWEKEA-----KDLT-----ELVORITYLQNPDKCSKARKLVGNI 214  
Db 949 KYLYKTH---VWIESYELPNKKYTPYDFENFRHVDVIOKKIQALQNEQN-----KLSINI 1000  
Qy 215 NK 216  
Db 1001 NR 1002

## RESULT 12

Q9SAA5 PRELIMINARY; PRT; 1025 AA.  
AC Q9SAA5;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE F25C20.13 protein.  
GN F25C20.13.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cnv. Columbia;  
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,  
RA Li J., Lee J.M., Kremenetskaia I., Lueros J., Ngan I., Liu A.,  
RA Gonzalez A., Altafi H., Araujo R., Chao Q., Conn L., Conway A.B.,  
RA Dunn P., Hansen N., Huizar L., Kim C., Palm C., Rowley D., Shinn P.,  
RA Walker M., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;

RT "Arabidopsis thaliana chromosome 1 BAC F25C20 sequence.";  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN=cnv. Columbia;  
RA Theologis A.;  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC007296; AAD30351.1;  
SQ SEQUENCE 1025 AA; 116411 MW; 08C952A2032BA1E4 CRC64;

Query Match 4.0%; Score 124; DB 10; Length 1025;  
Best Local Similarity 18.6%; Pred. No. 3.4;  
Matches 115; Conservative 79; Mismatches 184; Indels 240; Gaps 32;

Qy 5 TGSWRWIMLILFAWGTLLP-YI-----GCHLVDRNDHPDHSRRELSKILAKLERLKKQ 56  
Db 362 SGNW-WFAEVVVGALVIDWVADGPPKGAFLYDNGYQDFHALVPQKLPPELYWLEEE 420  
Qy 57 NEDLRRMAESLRPEGPI-----DQGTATGRVVRVLEEQIVKAKEQIENYKQARNDLG 109  
Db 421 NMIFRKLQEDRRLLKEEYMRAMKEMKTARLKAETKERTLKLLSKQVWV--YTEPLEIQAG 478  
Qy 110 KDHEIL---RRRTEGAKELWF--FLOSELKLLKLEGNELQORHADE-----ILLD 155  
Db 479 NPVTLYNPANTVILNGKPEVWFRGSRNWRTHRLGPLPPQKMEATDDESSHVKTAKVPLD 538  
Qy 156 LGHHSRIMTDLVYLSOTDGCAGWEKE-----AK 185  
Db 539 -----AYMDFVFSEKEDG-GIFDNKNGLDYHLVPVGGISKEPPLHIVHIAVEMAPIAK 591  
Qy 186 ---DLTELVOIRITYLQN-----PK-DCSK---ARKLVCN----- 213  
Db 592 VGGIGDVVTSLSRAVQELNHNVDIVFPKYDCIKHNFVKDLQFNRSYHWGGTEIKVWHGKV 651  
Qy 214 -----INKGCGYGCQ-----LHHVYVCPMIAYGTORTLILESQNRWYA 251  
Db 652 EGLSVYFLDPQNLQIFRGCVYGCADAGRFFGFFCHAALEFLQGGFHPD-ILHCHDWSSA 710  
Qy 252 TGGWETVFRPVSTCTDRSGLSGTGHS--GEVKKVQVWVLEPVLVDSLH----- 298  
Db 711 PVSW-----LFKDHYYQYGLIKTR-----IVFTIHNLEFGANAIGK 746  
Qy 299 -----PRPPYL-PLAVPEDLADRLRLRVHG-----DPAVW--WVSQVVKYLIRPQP 340  
Db 747 AMTFADKATTVSPYAKEVAGNSVISAHLYKFHGIINGIDPDIWDPYNDNFI-----PVP 801  
Qy 341 WLREI-----EETTKLGLFK---HPVIGVHVVRTDKVGTAAAPHPPIEYVWVVEEH 389  
Db 802 YTSENVVVEGKRAAKEELQNLRLGLKSADFPVVGIIITRLTHQKG-----IHLIKH 849  
Qy 390 --FOLLERRMKV-----DKKRVYLATDDPSLLKEAKTK 420  
Db 850 AIWETLERNGQVLLGSAPDPRIQNDFVNLANQLHSHGDRARLVITYDEPL-----SHLI 905  
Qy 421 YSNYEFTSDNSISWSAGL 438  
Db 906 YAGADFILVPSIPEPCGL 923

## RESULT 13

Q9UEG2 PRELIMINARY; PRT; 1285 AA.  
AC Q9UEG2;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Hypothetical protein KIAA0389.  
GN KIAA0389.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]



```

Qy 23 FYIGHLVRDNDPHDSRELSKILAKLERLKQONEDLRMAESLRIPGEPIDQGTATGR 82
Db 55 FEIEKRLSQSERLVTETRECONLRLELEKLNQVKVLTKEKLE-----TAQDR 105
Qy 83 VRVLEEQLVKAKEQIENYKQARDNLGKHDIILRRRIENGAKELWFFLOSELKXL--KKL 140
Db 106 NLGIQSOFTRAKELEAEKR---DLIRTNERLSQVEB-----YLTEDVKRLNEKJK 153
Qy 141 EGN---ELORHAEIILLDGHHERSINTDLYLSQTDGAGWEKEKAQDLTELVRORIT 196
Db 154 ESNTTKGELQKLDE-----LOADSVAVKYREKRELEQEKELHNNQNS 195
Qy 197 YLONPKDCSKARKLVCNINKG---CGYGCQLHHVYVCFMIAYGTOITLILESQNRYATG 253
Db 196 WLNTLTKTKTDDELLALGREKGNELILELKNLEN-----KKEEVLRLLEEQNGLKTS 246
Qy 254 GWETVFRPVSETCTDRSGLSTGHSGEVKQKNQVVELPI-----VDSLHPPPYLPL 306
Db 247 N-EHLQKHVEDLLTKLKEAQASMEKFFHNLNAHIKLSNLYKSAADDSEAKSNELTR 305
Qy 307 AVPE-----DLADLLRVHGDPAVWVVSQFVKYLIRPQWLEREIEETTKL 353
Db 306 AVDELHLKLKEAGEANKTIQDHLLOVESK-----DQMEKEMLEKIGKLEKELENANDLL 360
Qy 354 GFKHPVIGVHVVRTDKVQTE---AAFHP-----IEEYMHVVEEHFOLLER 395
Db 361 S-----ATKRGAILSEBELAAMSPTAAAVAKIVKPGMKLTLYNAYVETODQLLLE 412
Qy 396 RMKVDKRVYLATDDPSLLKEAKTKYSNYEFISDINSISWSAGLHNRVTENSLRGVILDIH 455
Db 413 --KQENKRIKYLDE--IVKEVEAK-----APILKQOREE-----443
Qy 456 FLSQADFLVCTFSQVCRVAYEIMOTLHPDASANFHS 492
Db 444 -YERAQKAVASLSAKLEQAMKEIQRLQEDTDKANKHS 479

```

Search completed: February 2, 2004, 08:43:08  
Job time : 37.5 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 08:39:39 ; Search time 38.5 Seconds  
(without alignments)  
2370.592 Million cell updates/sec

Title: US-09-971-773-24  
Perfect score: 3085  
Sequence: 1 MRAWTGSRWIMLILFAWGT.....YKREKIETVKYTPPEAEK 575

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq 19Jun03.\*

```

1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3085	100.0	575	23 ABG34136	Antibody productio
2	3024	98.0	575	23 ABG34135	Antibody productio
3	3003	97.3	575	18 AAW22125	Human alpha 1-6 fu
4	3003	97.3	575	23 ABB08405	Alpha1,6-fucosyl t
5	2910	94.3	575	18 AAW22124	Pig alpha 1-6 fuco
6	2658	86.2	515	22 AAG73884	Human colon cancer
7	1778	57.6	339	22 AAB75061	Human alpha 1-6 fu
8	1480.5	48.0	619	22 ABB59117	Drosophila melanog
9	1206	39.1	233	22 AAB75062	Human alpha 1-6 fu

10	432	14.0	82	22	ABG48804	Human liver peptid
11	432	14.0	82	22	ABB28797	Peptide #1448 enco
12	432	14.0	82	22	ABB33985	Peptide #1491 enco
13	432	14.0	82	22	ABB19422	Protein #1421 enco
14	432	14.0	82	22	AAM54747	Human brain expres
15	432	14.0	82	22	AAM67139	Human bone marrow
16	432	14.0	82	22	AAM15003	Peptide #1437 enco
17	432	14.0	82	22	AAM27440	Peptide #1477 enco
18	432	14.0	82	22	AAM02730	Peptide #1412 enco
19	432	14.0	82	22	ABG36801	Human peptid enco
20	426	13.8	81	22	AAM64859	Human brain expres
21	426	13.8	81	22	AAMJ7789	Peptide #1026 enc
22	131	4.2	25	18	AAM22131	Fragment #2 of hum
23	125	4.1	1025	23	ABB90967	Herbicideally activ
24	120.5	3.9	751	22	ABB62477	Drosophila melanog
25	120.5	3.9	3542	22	AAB62142	P. falciparum PCR3
26	118.5	3.8	976	24	ABF74709	Human SCP-1 protei
27	117	3.8	1456	22	ABB58673	Drosophila melanog
28	117	3.8	1521	21	AAG39235	Arabidopsis thalia
29	117	3.8	1528	21	AAG39234	Arabidopsis thalia
30	117	3.8	1552	21	AAG39233	Arabidopsis thalia
31	117	3.8	1703	21	AAG36714	Arabidopsis thalia
32	117	3.8	1710	21	AAG36713	Arabidopsis thalia
33	117	3.8	1744	21	AAG36712	Arabidopsis thalia
34	116.5	3.8	795	23	ABB77430	Human tumour marke
35	116	3.8	1235	24	ABR47540	Breast cancer asso
36	115.5	3.7	931	22	AAM79504	Human protein SEQ
37	115.5	3.7	930	22	AAM78520	Human protein sequ
38	114.5	3.7	612	22	AAB95546	Human homologue of
39	114.5	3.7	888	23	AAU83013	Human 160kD mediat
40	114.5	3.7	1427	12	AAI10534	Osteoclast stimula
41	114	3.7	214	18	AAW09037	Osteoclast stimula
42	114	3.7	214	18	AAW12706	Osteoclast stimula
43	114	3.7	229	17	AAW05403	Human clone 5 prot
44	113.5	3.7	976	22	AAG66581	Human SCP-1 muclein
45	113	3.7	26	18	AAW22126	Fragment #1 of Pig

#### ALIGNMENTS

#### RESULT 1

ABG34136  
ID ABG34136 standard; Protein; 575 AA.  
AC ABG34136;  
XX  
XX  
15-JUL-2002 (first entry)  
DT  
XX  
XX  
Antibody production method related protein #1.  
DE  
XX  
Antibody production; cytostatic; immunomodulator; vasotropic; virucide;  
KW  
antibacterial; antiinflammatory; antiallergic; allergy; inflammation;  
KW  
autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO;  
KW  
tumour; circulatory disease; infection.  
XX  
XX  
Mus musculus.  
OS  
XX  
WO200231140-A1.  
PN  
XX  
18-APR-2002.  
PD  
XX  
05-OCT-2001; 2001WO-JF08804.  
PF  
XX  
06-OCT-2000; 2000JP-0308526.  
PR  
XX  
(KYOW ) KYOWA HAKKO KCGYO KK.  
PA  
XX  
Kanda Y, Satoh M, Nakamura K, Uchida K, Shinkawa T, Yamane N;  
PI  
Hosaka E, Yamano K, Yamasaki M, Hanai N;  
XX  
WPI; 2002-340182/37.  
DR  
XX

PT Cells producing antibody compositions including antibody fragments and  
PT fusion proteins with Fc domain of antibody, useful for prevention or  
PT treatment of cancer, immune diseases, circulatory diseases and  
PT infections  
XX  
XX  
PS Claim 33; Page 21-23; 314pp; Japanese.  
XX This invention relates to novel method for antibody production  
CC comprising a Chinese hamster ovarian tissue-originated (CHO) cell  
CC transferred with a gene encoding an antibody molecule for producing a  
CC composition comprising an antibody molecule with an Fc domain bonded  
CC to the N-glycoside linkage complex sugar chain. The produced antibody  
CC compositions are drugs for prevention or treatment of diseases  
CC accompanying tumour, allergy or inflammation, autoimmune diseases,  
CC circulatory diseases, and viral and bacterial infections. The  
CC antibodies can be stably produced using the method of the invention  
CC with high binding activity and potency thus leading to high safety and  
CC reduced side effects when applied alone or in combination with other  
CC drugs for therapy. The present sequence represents a protein  
CC molecule used in the method of the invention.  
XX  
XX  
SQ Sequence 575 AA;  
Query Match 100.0%; Score 3085; DB 23; Length 575;  
Best Local Similarity 100.0%; Pred. No. 3.1e-279;  
Matches 575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MRAWTGSRWIMLILFAWGTLFYIGGHLVRNDHPDHSRELSKILAKLERLKQNEEDL 60  
DB 1 MRAWTGSRWIMLILFAWGTLFYIGGHLVRNDHPDHSRELSKILAKLERLKQNEEDL 60  
QY 61 RMAESLRIPGPIDQGTATGRVRLVEQLVKAKEIQENYKKQARNGLGKHDEILRRRIE 120  
DB 61 RMAESLRIPGPIDQGTATGRVRLVEQLVKAKEIQENYKKQARNGLGKHDEILRRRIE 120  
QY 121 NGAKELWFFLQSELKHLKLEGNELQORHADEILDLGHHERSIMTDLVYLSQTDGAGDWR 180  
DB 121 NGAKELWFFLQSELKHLKLEGNELQORHADEILDLGHHERSIMTDLVYLSQTDGAGDWR 180  
QY 181 EKEAKDLTELVRRTIYLNQPKDCSKARKLVNCINKGCGYQLHHVYCFMAYGTORT 240  
DB 181 EKEAKDLTELVRRTIYLNQPKDCSKARKLVNCINKGCGYQLHHVYCFMAYGTORT 240  
QY 241 LILESQNRWYATGWTVPFRVPSVETCTDRSGLSGTGHSGEVNDKNIQVVELPIVDSLHPR 300  
DB 241 LILESQNRWYATGWTVPFRVPSVETCTDRSGLSGTGHSGEVNDKNIQVVELPIVDSLHPR 300  
QY 301 PPYLPVAPVEDLADRLRLRVHGDPAVWVSQFVKYLIRPQWLEKEIEEATKLGPKHPVI 360  
DB 301 PPYLPVAPVEDLADRLRLRVHGDPAVWVSQFVKYLIRPQWLEKEIEEATKLGPKHPVI 360  
QY 361 GVHVRTDKVGEAAFPHEEYVWVHEEHFQLLARRMQVKKRVYLATDDPTLLKEATK 420  
DB 361 GVHVRTDKVGEAAFPHEEYVWVHEEHFQLLARRMQVKKRVYLATDDPTLLKEATK 420  
QY 421 YSNYEFISDNSISWSAGLHNRYTENSRLGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
DB 421 YSNYEFISDNSISWSAGLHNRYTENSRLGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
QY 481 TLHPDASANFHLDDIYYPGGQNAHQIAVYPKHPRTBEEIPMEPGDIIIGVAGNHWGYS 540  
DB 481 TLHPDASANFHLDDIYYPGGQNAHQIAVYPKHPRTBEEIPMEPGDIIIGVAGNHWGYS 540  
QY 541 KGINKRLGKTGLYPSYKVRKEITETVKYPTYPEAK 575  
DB 541 KGINKRLGKTGLYPSYKVRKEITETVKYPTYPEAK 575

RESULT 2

ABG34135

ID ABG34135 standard; Protein; 575 AA.

XX

AC ABG34135;

XX  
DT 15-JUL-2002 (first entry)  
DE Antibody production method related protein #1.  
XX  
XX  
KW Antibody production; cytostatic; immunomodulator; vasotropic; virucide;  
KW antibacterial; antiinflammatory; antiallergic; allergy; inflammation;  
KW autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO;  
XX tumour; circulatory disease; infection.  
OS Cricetus griseus.  
XX WO200231140-A1.  
XX  
XX 18-APR-2002.  
XX  
XX 05-OCT-2001; 2001WO-JP08804.  
XX  
XX 06-OCT-2000; 2000JP-0308526.  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
XX Kanda Y, Satoh M, Nakamura K, Uchida K, Shinkawa T, Yamane N;  
PI Hosaka E, Yamano K, Yamasaki M, Hanai N;  
XX  
XX WPI; 2002-340182/37.  
XX  
XX Cells producing antibody compositions including antibody fragments and  
PT fusion proteins with Fc domain of antibody, useful for prevention or  
PT treatment of cancer, immune diseases, circulatory diseases and  
PT infections  
XX  
XX  
PS Claim 23; Page 18-21; 314pp; Japanese.  
XX  
XX This invention relates to novel method for antibody production  
CC comprising a Chinese hamster ovarian tissue-originated (CHO) cell  
CC transferred with a gene encoding an antibody molecule for producing a  
CC composition comprising an antibody molecule with an Fc domain bonded  
CC to the N-glycoside linkage complex sugar chain. The produced antibody  
CC compositions are drugs for prevention or treatment of diseases  
CC accompanying tumour, allergy or inflammation, autoimmune diseases,  
CC circulatory diseases, and viral and bacterial infections. The  
CC antibodies can be stably produced using the method of the invention  
CC with high binding activity and potency thus leading to high safety and  
CC reduced side effects when applied alone or in combination with other  
CC drugs for therapy. The present sequence represents a protein  
CC molecule used in the method of the invention.  
XX  
XX  
SQ Sequence 575 AA;  
Query Match 98.0%; Score 3024; DB 23; Length 575;  
Best Local Similarity 97.6%; Pred. No. 1.6e-273;  
Matches 561; Conservative 9; Mismatches 5; Indels 0; Gaps 0;  
QY 1 MRAWTGSRWIMLILFAWGTLFYIGGHLVRNDHPDHSRELSKILAKLERLKQNEEDL 60  
DB 1 MRAWTGSRWIMLILFAWGTLFYIGGHLVRNDHPDHSRELSKILAKLERLKQNEEDL 60  
QY 61 RMAESLRIPGPIDQGTATGRVRLVEQLVKAKEIQENYKKQARNGLGKHDEILRRRIE 120  
DB 61 RMAESLRIPGPIDQGTATGRVRLVEQLVKAKEIQENYKKQARNGLGKHDEILRRRIE 120  
QY 121 NGAKELWFFLQSELKHLKLEGNELQORHADEILDLGHHERSIMTDLVYLSQTDGAGDWR 180  
DB 121 NGAKELWFFLQSELKHLKLEGNELQORHADEILDLGHHERSIMTDLVYLSQTDGAGDWR 180  
QY 181 EKEAKDLTELVRRTIYLNQPKDCSKARKLVNCINKGCGYQLHHVYCFMAYGTORT 240  
DB 181 EKEAKDLTELVRRTIYLNQPKDCSKARKLVNCINKGCGYQLHHVYCFMAYGTORT 240  
QY 241 LILESQNRWYATGWTVPFRVPSVETCTDRSGLSGTGHSGEVNDKNIQVVELPIVDSLHPR 300  
DB 241 LILESQNRWYATGWTVPFRVPSVETCTDRSGLSGTGHSGEVNDKNIQVVELPIVDSLHPR 300

QY 301 PPYLPLAVPEDLADRLRVHGDPAVWVVSQFVKYLIRPQWLEKEIEEATKLGFKHPVI 360  
 DB 301 PPYLPLAVPEDLADRLRVHGDPAVWVVSQFVKYLIRPQWLEKEIEEATKLGFKHPVI 360  
 QY 361 GVHVRTDKVGEAAAFHPIEYVMVHVEBHFOQLARRMQVDKRVYLAATDDPTLLKEATK 420  
 DB 361 GVHVRTDKVGEAAAFHPIEYVMVHVEBHFOQLARRMQVDKRVYLAATDDPTLLKEATK 420  
 QY 421 YSNYEFISDINSISAGLHNYTENSRLGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
 DB 421 YSNYEFISDINSISAGLHNYTENSRLGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
 QY 481 TLHPDASANFSLDDIYVFGGONAHNOIAVYVPHKPRTEEEIPMBEFGDIIGVAGNHWGYS 540  
 DB 481 TLHPDASANFSLDDIYVFGGONAHNOIAVYVPHKPRTEEEIPMBEFGDIIGVAGNHWGYS 540  
 QY 541 KGINKLGTGLYPSYKVKREKIETVKYPTYPEAEK 575  
 DB 541 KGVNRKLGKGLYPSYKVKREKIETVKYPTYPEAEK 575  
 RESULT 3  
 AAW22125  
 ID AAW22125 standard; Protein; 575 AA.  
 XX AC AAW22125;  
 XX DT 05-MAR-1998 (first entry)  
 XX DE Human alpha 1-6 fucosyltransferase.  
 XX KW Alpha 1-6 fucosyltransferase; enzyme; pig; human; fucose transfer;  
 KW guanosine diphosphate; sugar chain synthesis; modification; antibody;  
 KW GlnAc; cancer diagnosis.  
 XX OS Homo sapiens.  
 XX PN WO9727303-A1.  
 XX PD 31-JUL-1997.  
 XX PF 23-JAN-1997; 97WO-JP00171.  
 XX PR 22-JUL-1996; 96JP-0192260.  
 PR 24-JAN-1996; 96JP-0010365.  
 PR 21-JUN-1996; 96JP-0161648.  
 PR 24-JUN-1996; 96JP-0162813.  
 XX PA (TOYM ) TOYO BOSEKI KK.  
 XX PI Shiba T, Taniguchi N, Uozumi N, Yanagidani S;  
 XX DR WPI; 1997-393690/36.  
 DR N-PSDB; AAT76574.  
 XX Human or pig alpha 1-6 fucosyltransferase and DNA encoding it - for  
 PT synthesis and modification of sugar chains and used as an antigen  
 PT for production of diagnostic antibodies  
 XX Claim 17; Page 39-43; 61pp; Japanese.  
 XX AAW22124 and AAW22125 represent the pig and human alpha 1-6  
 CC fucosyltransferases of the invention, respectively. The enzyme transfers  
 CC fucose from guanosine diphosphate to the 6-hydroxyl group of the GlnAc  
 CC nearest to R in the receptor molecule: (GlnAc-beta 1-2Man-alpha 1-6)  
 CC (GlnAc-beta 1-2Man-alpha 1-3)Man-beta 1-4GlcNAc-beta 1-4GlcNAc-R to give  
 CC (GlnAc-beta 1-2Man-alpha 1-6)(GlcNAc-beta 1-2Man-alpha 1-3)Man-beta  
 CC 1-4GlcNAc-beta 1-4(Fuc-alpha 1-6)(GlcNAc-R. It has an optimum pH of about  
 CC 7.0 (pig) or 7.5 (human), and is stable over the pH range 4-10 after 5  
 CC hours at 4 degrees C. The optimum working temperature of the  
 CC alpha 1-6 fucosyltransferases is 30-37 degrees C. A bivalent metal is  
 CC not required for activity of the enzyme, and the enzyme is not inhibited

CC in the presence of 5 mM EDTA. The enzyme is useful in the synthesis and  
 CC modification of sugar chains, and as antigen for the production of  
 CC antibodies recognising the enzyme. The antibodies can be used for the  
 CC diagnosis of cancer and other diseases.  
 XX SQ Sequence 575 AA;  
 Query Match 97.3%; Score 3003; DB 18; Length 575;  
 Best Local Similarity 96.5%; Pred. No. 1.4e-271;  
 Matches 555; Conservative 11; Mismatches 9; Indels 0; Gaps 0;  
 QY 1 MRWTGSRWRIMLILFWAGTLLFFVIGHLVRDNDHPDHSRELSKILAKLERLQONEDL 60  
 DB 1 MRPTGSRWRIMLILFWAGTLLFFVIGHLVRDNDHPDHSRELSKILAKLERLQONEDL 60  
 QY 61 RRMASRLRIPGPIIDQGTATGRVRLVEQLVKAKEQIENYKQARNGLKGDKHEILRRRIE 120  
 DB 61 RRMASRLRIPGPIIDQGTATGRVRLVEQLVKAKEQIENYKQARNGLKGDKHEILRRRIE 120  
 QY 121 NGAKELWFFLQSELKHLKLEGNELQRADEFLDLGHHERSINTDLYLSQTDGAGDWR 180  
 DB 121 NGAKELWFFLQSELKHLKLEGNELQRADEFLDLGHHERSINTDLYLSQTDGAGDWR 180  
 QY 181 EKEAKDLTELVRITTYLQNPDKCSKARKLVNINCKGCGYCOLHHVYCFMIAYGTOPT 240  
 DB 181 EKEAKDLTELVRITTYLQNPDKCSKARKLVNINCKGCGYCOLHHVYCFMIAYGTOPT 240  
 QY 241 LILESQNRVYATGWTGTVFRPVSETCTDRSGLSTGHSGEVNDKNOVVELPIVDSLHPR 300  
 DB 241 LILESQNRVYATGWTGTVFRPVSETCTDRSGLSTGHSGEVNDKNOVVELPIVDSLHPR 300  
 QY 301 PPYLPLAVPEDLADRLRVHGDPAVWVVSQFVKYLIRPQWLEKEIEEATKLGFKHPVI 360  
 DB 301 PPYLPLAVPEDLADRLRVHGDPAVWVVSQFVKYLIRPQWLEKEIEEATKLGFKHPVI 360  
 QY 361 GVHVRTDKVGEAAAFHPIEYVMVHVEBHFOQLARRMQVDKRVYLAATDDPTLLKEATK 420  
 DB 361 GVHVRTDKVGEAAAFHPIEYVMVHVEBHFOQLARRMQVDKRVYLAATDDPTLLKEATK 420  
 QY 421 YSNYEFISDINSISAGLHNYTENSRLGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
 DB 421 YSNYEFISDINSISAGLHNYTENSRLGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
 QY 481 TLHPDASANFSLDDIYVFGGONAHNOIAVYVPHKPRTEEEIPMBEFGDIIGVAGNHWGYS 540  
 DB 481 TLHPDASANFSLDDIYVFGGONAHNOIAVYVPHKPRTEEEIPMBEFGDIIGVAGNHWGYS 540  
 QY 541 KGINKLGTGLYPSYKVKREKIETVKYPTYPEAEK 575  
 DB 541 KGVNRKLGKGLYPSYKVKREKIETVKYPTYPEAEK 575  
 RESULT 4  
 ABB08405  
 ID ABB08405 standard; Protein; 575 AA.  
 XX AC ABB08405;  
 XX DT 07-MAY-2002 (first entry)  
 XX DE Alpha1,6-fucosyl transferase amino acid sequence.  
 XX KW Plant; glycoprotein; alpha1,6-fucosyl transferase; alpha1,6-Ft;  
 KW enzyme.  
 XX OS Homo sapiens.  
 XX PN JP2001333787-A.  
 XX PD 04-DEC-2001.  
 XX PF 06-MAR-2001; 2001JP-0062704.  
 XX

```

PR 22-MAR-2000; 2000JP-0081059.
XX
XX (TANI/) TANIGUCHI N.
PA (SEKI/) SEKI T.
PA (FUJI/) FUJIYAMA K.
XX
XX WPI; 2002-159816/21.
DR N-PSDB; ABA98809.
XX
XX A plant cell with an animal type sugar chain adding function, for the
PT preparation of a glycoprotein with an animal type sugar chain -
XX
XX Example 1; Page 24-26; 38pp; Japanese.
XX
XX The invention relates to a plant cell with an animal type sugar chain
CC adding function, created by transforming a tobacco cell with a gene
CC encoding an enzyme derived from an animal which can transfer a fucose
CC residue to the reductive end acetylglucosamine residue of a sugar chain.
CC The gene that is introduced into the plant cell encodes the enzyme
CC alpha1,6-fucosyl transferase. The method of the invention is useful for
CC the preparation of a glycoprotein having animal type sugar chain. The
CC current sequence represents alpha1,6-fucosyl transferase amino acid
CC sequence.
XX
XX Sequence 575 AA;
SQ
Query Match 97.3%; Score 3003; DB 23; Length 575;
Best Local Similarity 96.5%; Pred. No. 1.4e-271;
Matches 555; Conservative 11; Mismatches 9; Indels 0; Gaps 0;
QY 1 MRWGTSGWRWIMLILFANGTLLFYIGGHLVRDNDHPDSSRELSKILAKLERLKQONEDL 60
Db 1 MRWGTSGWRWIMLILFANGTLLFYIGGHLVRDNDHPDSSRELSKILAKLERLKQONEDL 60
QY 61 RMAESLRIPGPIDQGTATGRVLEOLVKAEQIENYKQARNGLGKDHILRRRIE 120
Db 61 RMAESLRIPGPIDQGTATGRVLEOLVKAEQIENYKQARNGLGKDHILRRRIE 120
QY 121 NGAKELWFFLQSELKHLKLEGNELQRAHDEFLDLGHHSRIMTDLVYLSQTDGAGDWR 180
Db 121 NGAKELWFFLQSELKHLKLEGNELQRAHDEFLDLGHHSRIMTDLVYLSQTDGAGDWR 180
QY 181 EKEAKDLTELVRITTYLQNPDKCSKARKLVNINKGCGYGCQLHHVYCFMIAVGTQRT 240
Db 181 EKEAKDLTELVRITTYLQNPDKCSKARKLVNINKGCGYGCQLHHVYCFMIAVGTQRT 240
QY 241 LILESQNRWYATGGWETVPRVSETCTDRSGISTGHWSGEVNDKNIQVVELPIVDSLHPR 300
Db 241 LILESQNRWYATGGWETVPRVSETCTDRSGISTGHWSGEVNDKNIQVVELPIVDSLHPR 300
QY 301 PPYLPLAVPEDLADRLRLVHGDPVWVWSQFVKYLIRPQWLEKEIEBEATKLGFKHPVI 360
Db 301 PPYLPLAVPEDLADRLRLVHGDPVWVWSQFVKYLIRPQWLEKEIEBEATKLGFKHPVI 360
QY 361 GVHVRTDKVGTGAAPHPIEYMWVHEHFQLLARMQVDKRVYLATDPTLLKEATK 420
Db 361 GVHVRTDKVGTGAAPHPIEYMWVHEHFQLLARMQVDKRVYLATDPTLLKEATK 420
QY 421 YSNYEFISDNSISWSAGLHNRYTENSRLGVLDIHFLSQADFLVCTFSQVCRVAYEIMQ 480
Db 421 YSNYEFISDNSISWSAGLHNRYTENSRLGVLDIHFLSQADFLVCTFSQVCRVAYEIMQ 480
QY 481 TLHPDASANFHLDDIYYPGGQANNOIAVYPHKPRTTEEIPMEPGDIIIGVAGNHWDCYS 540
Db 481 TLHPDASANFHLDDIYYPGGQANNOIAVYPHKPRTTEEIPMEPGDIIIGVAGNHWDCYS 540
QY 541 KGINRKLGTGLYPSYKVKREKTIETVKYPTYPEAK 575
Db 541 KGVNRKLGTGLYPSYKVKREKTIETVKYPTYPEAK 575
RESULT 5
AAW22124

```

AAW22124 standard; Protein; 575 AA.  
 AAW22124;  
 05-MAR-1998 (first entry)  
 Pig alpha 1-6 fucosyltransferase.  
 Alpha 1-6 fucosyltransferase; enzyme; pig; human; fucose transfer;  
 guanosine diphosphate; sugar chain synthesis; modification; antibody;  
 GlcNAc; cancer diagnosis.  
 Sus scrofa.  
 WO9727303-A1.  
 31-JUL-1997.  
 23-JAN-1997; 97WO-JP00171.  
 22-JUL-1996; 96JP-0192260.  
 24-JAN-1996; 96JP-0010365.  
 21-JUN-1996; 96JP-0161648.  
 24-JUN-1996; 96JP-0162813.  
 (TOYM) TOYO BOSEKI KK.  
 Shiba T, Taniguchi N, Uozumi N, Yanagidani S;  
 WPI; 1997-393690/36.  
 N-PSDB; AAT76573.  
 Human or pig alpha 1-6 fucosyltransferase and DNA encoding it - for  
 synthesis and modification of sugar chains and used as an antigen  
 for production of diagnostic antibodies  
 Claim 4; Page 30-34; 61pp; Japanese.  
 AAW22124 and AAW22125 represent the pig and human alpha 1-6  
 fucosyltransferases of the invention, respectively. The enzyme transfers  
 fucose from guanosine diphosphate to the 6-hydroxyl group of the GlcNAc  
 nearest to R in the receptor molecule: (GlcNAc-beta 1-2-Man-alpha 1-6)  
 (GlcNAc-beta 1-2-Man-alpha 1-3)-Man-beta 1-4-GlcNAc-beta 1-4-GlcNAc-R to give  
 (GlcNAc-beta 1-2-Man-alpha 1-6) (GlcNAc-beta 1-2-Man-alpha 1-3)-Man-beta  
 1-4-GlcNAc-beta 1-4 (Fuc-alpha 1-6) GlcNAc-R. It has an optimum pH of about  
 7.0 (pig) or 7.5 (human), and is stable over the pH range 4-10 after 5  
 hours at 4 degrees C. The optimum working temperature of the  
 alpha 1-6 fucosyltransferases is 30-37 degrees C. A bivalent metal is  
 not required for activity of the enzyme, and the enzyme is not inhibited  
 in the presence of 5 mM EDTA. The enzyme is useful in the synthesis and  
 modification of sugar chains, and as antigen for the production of  
 antibodies recognising the enzyme. The antibodies can be used for the  
 diagnosis of cancer and other diseases.

Query Match 94.3%; Score 2910; DB 18; Length 575;  
 Best Local Similarity 93.6%; Pred. No. 7.1e-263;  
 Matches 538; Conservative 14; Mismatches 23; Indels 0; Gaps 0;  
 QY 1 MRWGTSGWRWIMLILFANGTLLFYIGGHLVRDNDHPDSSRELSKILAKLERLKQONEDL 60  
 Db 1 MRWGTSGWRWIMLILFANGTLLFYIGGHLVRDNDHPDSSRELSKILAKLERLKQONEDL 60  
 QY 61 RMAESLRIPGPIDQGTATGRVLEOLVKAEQIENYKQARNGLGKDHILRRRIE 120  
 Db 61 RMAESLRIPGPIDQGTATGRVLEOLVKAEQIENYKQARNGLGKDHILRRRIE 120  
 QY 121 NGAKELWFFLQSELKHLKLEGNELQRAHDEFLDLGHHSRIMTDLVYLSQTDGAGDWR 180  
 Db 121 NGAKELWFFLQSELKHLKLEGNELQRAHDEFLDLGHHSRIMTDLVYLSQTDGAGDWR 180  
 QY 181 EKEAKDLTELVRITTYLQNPDKCSKARKLVNINKGCGYGCQLHHVYCFMIAVGTQRT 240

Db 181 EKEAKDLTELVRITTYLQNPDKCSKAKKLVNKNKGGYGCQLHHVYVCFMIAVGTQRT 240  
Qy 241 LILESQWRYATGGWETFRPVSCTDRSGLSTGHWSGEVNDKNIQVVELPIVDSLHPR 300  
Db 241 LALESHNNRYATGGWETFRPVSCTDRSGSSTGHWSGEVNDKNIQVVELPIVDSLHPR 300  
Qy 301 PYPPLAVPDLADRLLRVHGDPAVWVVSQFVKYLIRPQWLEKEIEEATKKGPKHPVI 360  
Db 301 PYPPLAVPDLADRLVRVHGDPAVWVVSQFVKYLIRPQWLEKEIEEATKKGPKHPVI 360  
Qy 361 GVHVRTDKVTEAAFHPIEYVHVEHFFQLLARRMQVKKRVYLATDDPTLLKEATK 420  
Db 361 GVHVRTDKVTEAAFHPIEYVHVEHFFQLLARRMQVKKRVYLATDDPTLLKEATK 420  
Qy 421 YSNYEFISDNTSWAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSQVCRVAYETMQ 480  
Db 421 YSNYEFISDNTSWAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSQVCRVAYETMQ 480  
Qy 481 TLHPDASANFSLDDIYFGGQNAHQIAVYPHKPRTEEEIPMEPGDIIGVAGNHWDCYS 540  
Db 481 ALHPDASANFSLDDIYFGGQNAHQIAVYPHKPRTEEEIPMEPGDIIGVAGNHWDCYS 540  
Qy 541 KGINKLGTGLYPSYKREKTIETVKYPTYPEAK 575  
Db 541 KGVNKLGTGLYPSYKREKTIETVKYPTYPEADK 575

RESULT 6  
AAG73884  
ID AAG73884 standard; Protein; 515 AA.  
XX AC AAG73884;  
XX DT 03-SEP-2001 (first entry)  
XX DE Human colon cancer antigen protein SEQ ID NO:4648.  
XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
XX KW Colorectal carcinoma; chromosome 14.  
XX OS Homo sapiens.  
XX PN WO200122920-A2.  
XX PD 05-APR-2001.  
XX PF 28-SEP-2000; 2000WO-US26524.  
XX PR 29-SEP-1999; 99US-0157137.  
XX PR 03-NOV-1999; 99US-0163280.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;  
XX DR WPI: 2001-235357/24.  
XX DR N-PSDB; AAH33315.  
XX PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
XX PT useful for preventing, diagnosing and/or treating colorectal cancers -  
XX PS Claim 11; Page 6451-6453; 9803pp; English.  
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
XX cancer-associated nucleic acid molecules (N) and proteins (P), where  
XX the proteins are collectively known as colon cancer antigens. The colon  
XX cancer antigens have cytostatic activity and can be used in gene  
XX therapy and vaccine production. N and P may be used in the prevention,  
XX diagnosis and treatment of diseases associated with inappropriate P  
XX expression. For example, N and P may be used to treat disorders  
XX associated with decreased expression by rectifying mutations or deletions  
XX in a patient's genome that affect the activity of P by expressing

CC inactive proteins or to supplement the patients own production of P.  
CC Additionally, N may be used to produce the colon cancer-associated Ps,  
CC by inserting the nucleic acids into a host cell and culturing the cell  
CC to express the proteins. N and P can be used in the prevention, diagnosis  
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
CC and AAB77789 represent sequences used in the exemplification of the  
CC present invention.  
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
CC missing at time of publication, meaning no sequences are present for  
CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
XX

Qy 65 ESLRIPGPDIDQ7ATGRVRLVLEQVLKAKQIENYKQARGLGKHHEILRRRIENGAK 124  
Db 5 KSLRIPGPDIDQ7PAIGRVRLVLEQVLKAKQIENYKQARGLGKHHEILRRRIENGAK 64  
Qy 125 ELWFFLOSELKKGKLEGNELQORHAEITLLDLGHHSIMTDLYLSOTDCAGDWREKEA 184  
Db 65 ELWFFLOSELKKGKLEGNELQORHAEITLLDLGHHSIMTDLYLSOTDCAGDWREKEA 124  
Qy 185 KDLTELVRITTYLQNPDKCSKAKKLVNKNKGGYGCQLHHVYVCFMIAVGTQRTTILE 244  
Db 125 KDLTELVRITTYLQNPDKCSKAKKLVNKNKGGYGCQLHHVYVCFMIAVGTQRTTILE 184  
Qy 245 SONRYATGWEVFRPVSCTDRSGLSTGHWSGEVNDKNIQVVELPIVDSLHPRPPL 304  
Db 185 SONRYATGWEVFRPVSCTDRSGISTGHWSGEVNDKNIQVVELPIVDSLHPRPPL 244  
Qy 305 PLAVPEDLADRLVRVHGDPAVWVVSQFVKYLIRPQWLEKEIEEATKKGPKHPVIGVHV 364  
Db 245 PLAVPEDLADRLVRVHGDPAVWVVSQFVKYLIRPQWLEKEIEEATKKGPKHPVIGVHV 304  
Qy 365 RRTDKVGTAAAFHPIEYVHVEHFFQLLARRMQVKKRVYLATDDPTLLKEATKYSNY 424  
Db 305 RRTDKVGTAAAFHPIEYVHVEHFFQLLARRMQVKKRVYLATDDPTLLKEATKYSNY 364  
Qy 425 EFISDNTSWAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSQVCRVAYETMQTLHP 484  
Db 365 EFISDNTSWAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSQVCRVAYETMQTLHP 424  
Qy 485 DASANFSLDDIYFGGQNAHQIAVYPHKPRTEEEIPMEPGDIIGVAGNHWDCYSKGIN 544  
Db 425 DASANFSLDDIYFGGQNAHQIAVYPHKPRTEEEIPMEPGDIIGVAGNHWDCYSKGIN 544  
Qy 545 RKLKGTGLYPSYKREKTIETVKYPTYPEAK 575  
Db 485 RKLKGTGLYPSYKREKTIETVKYPTYPEAK 515

RESULT 7  
AAB75061  
ID AAB75061 standard; Protein; 339 AA.  
XX AC AAB75061;  
XX DT 20-JUL-2001 (first entry)  
XX DE Human alpha 1-6 fucosyltransferase protein 237-575 SEQ ID NO:3.  
XX KW Human; alpha 1-6 fucosyltransferase; alpha 1-6 fucT; antibody;  
XX KW alpha 1-6 fucose transferase; anti-human alpha 1-6 fucT; immunoassay.  
XX OS Homo sapiens.  
XX PN JP20001011097-A.  
XX PD 16-JAN-2001.  
XX

```
PF 29-JUN-1999; 99JP-0183569.
XX
XX 29-JUN-1999; 99JP-0183569.
XX
XX (FJRE ) FUJIREBIO KK.
XX
XX WPI; 2001-275926/29.
XX
XX N-PSDB; AAF87952.
XX
XX Novel anti-human alpha1-6 fucose transferase antibody useful for
XX immunoassay -
XX
XX Example 2; Page 8-9; 1lpp; Japanese.
XX
XX The present invention describes an anti-human alpha 1-6
XX fucosyltransferase (alpha 1-6 fuct, also called alpha 1-6 fucose
XX transferase) antibody, also described are: (1) a hybridoma producing
XX the above monoclonal antibody; (2) an immunoassay for detecting human
XX alpha 1-6 fuct by using the above antibody or its antibody fragment;
XX and (3) a reagent used for the above immunoassay. The anti-human
XX alpha 1-6 fuct antibody can be used for immunoassay. The present
XX sequence represents the human alpha 1-6 fuct protein of residues 237
XX to 575, which is used in an example from the present invention.
XX
XX Sequence 339 AA;
SQ
Query Match 57.6%; Score 1778; DB 22; Length 339;
Best Local Similarity 96.2%; Pred. No. 2.3e-157;
Matches 326; Conservative 9; Mismatches 4; Indels 0; Gaps 0;
QY 237 TORTLLESQNRWYATGGWETVFRPVSETCTDRSLGSLTGHWSGEVNDKNIQVVELPIVDS 296
DB 1 TORTLLESQNRWYATGGWETVFRPVSETCTDRSLGSLTGHWSGEVNDKNIQVVELPIVDS 60
QY 297 LHRPPYLPLAVPEDLADRLVHGDPVAVWYVSQVKYLIRPQWLEKEIEEATKLGFK 356
DB 61 LHRPPYLPLAVPEDLADRLVHGDPVAVWYVSQVKYLIRPQWLEKEIEEATKLGFK 120
QY 357 HPVIGVHVARTDKVGTGAFAFPIEEYMHVVEEHFOLLARRMOVDKRVYLATDDPTLLKE 416
DB 121 HPVIGVHVARTDKVGTGAFAFPIEEYMHVVEEHFOLLARRMOVDKRVYLATDDPTLLKE 180
QY 417 ATKYSNYEFDISNLSWSAGLHNRVTENSLRGVILDHFLSQADFLVCTFSSQVCRVAY 476
DB 181 ATKYPNYEFDISNLSWSAGLHNRVTENSLRGVILDHFLSQADFLVCTFSSQVCRVAY 240
QY 477 EIMQTLHPDASANFHSLLDDIYVFGGQNAHNOIAIYAHQPTADEIPMEPGDIIIGVAGNHW 536
DB 241 EIMQTLHPDASANFHSLLDDIYVFGGQNAHNOIAIYAHQPTADEIPMEPGDIIIGVAGNHW 300
QY 537 DGYSKGINRKLGTGLYPSYKVKREKTIETVKYPTYPEAEK 575
DB 301 DGYSKGVNRKLGTGLYPSYKVKREKTIETVKYPTYPEAEK 339
RESULT 8
ABB59117
ID ABB59117 standard; Protein; 619 AA.
XX
XX ABB59117;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 4143.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW
XX
XX pharmaceutical.
KW
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
```

```
XX 23-MAR-2001; 2001WO-US09231.
PF
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PMD, Myers EW;
PI
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL03220.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure; SEQ ID NO 4143; 2lpp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 619 AA;
SQ
Query Match 48.0%; Score 1480.5; DB 22; Length 619;
Best Local Similarity 46.5%; Pred. No. 4e-129;
Matches 285; Conservative 108; Mismatches 169; Indels 51; Gaps 8;
QY 3 AWTCSW-RWTMLILFANGTLLFYIGHLVRDN-----DHPDHSRSLSKILAKLERLKKQ 56
DB 10 ASANSWARALIFLVAWIGLVYVFWVKLTNTQGOQAAGESELNARRISQALQMLEHTQR 69
QY 57 NEDLRRMAESLRIPEGPTDQGTATGRVRLVEQLVKAQEIE----- 98
DB 70 NEELKQLIDELMSDQ--LDKQSAKMLVORLENDALNPKLAPEVAGPBPESFESAPADLR 127
QY 99 ---NYKKQARNGL-----GKDHEILRRRIENGAKELWFFLQSELSKCLK----- 138
DB 128 GWNVVAEGAPNDLEAGVPDHGEFEPSELEYEFTRRIQTNGEINWFFSELGKVRKAVAA 187
QY 139 -HLEGNELQRHADIILLDLGHHSRIMTDLYLVSQTDGAGDWREKAKDLTELVRRTY 197
DB 188 GHASA-DLEESINQVLLQGAHKRSLSDMRMRQSDGYEARHKEARDLSDLVQRRLHH 246
QY 198 LQNPDCSKARKLVNINKKCGYCOLHHVYVCEWIAVGTORTLILSQNRWYATGGWET 257
DB 247 LQNPESDQNAKLVCKLNKCGYCOLHHVYVCFIVAYATERILLKSRGWRHYKHGGWEE 306
QY 258 VFRPVSETCTDRSGLSTGHWSGEVNDKNIQVVELPIVDSLHPRPPYLPVAVPEDLADRL 317
DB 307 VFPVSNVSCHDAGTANTYNNWPK---PNTQVLVLPIDSLMPPRPYLPVAVPEDLAPLK 363
QY 318 RVHGDPAVWVVSQVKYLIRPQWLEKEIEEATKLGFKHPIGVHVHVARTDKVGTGAFAFH 377
DB 364 RLHGDPVWVWVGQFLKYLRLRPQTTTRDFTLSGMRNLGWERPIVGHVHVARTDKVGTGAACH 423
QY 378 PIEEYMHVVEEHFOLLARRMOVDKRVYLATDDPTLLKEAKTKYSNYEFDISNLSWSAG 437
DB 424 SVEEYMTVYEDYRTLEVNGSTVARRIFLASDDAQVIEEARRKYPOQIIGDPEVARNAS 483
QY 438 LHNRYTENSRLGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQTLHPDASANFHSLLDDIY 497
DB 484 VSTRYTDALNGIILDIHLLSMSDLHVLCTFSSQVCRVAYEIMQTMYPDAHRFKSLDDIY 543
```

QY 498 YFGQNAHNOIAVPHKPTBEEIPMEPGDIIIGVAGNHWGDSKGINRKLKGTGLYPSYK 557  
DB 544 YFGQNAHNRVIAHKRTHEDLQLRVGDLVSVAGNHWGDSKGNKTRTNQGGLPSPFK 603  
QY 558 VREKIEYKTYPT 570  
DB 604 VEKVDYAKPLV 616

RESULT 9  
AAB75062  
ID AAB75062 standard; Protein; 233 AA.  
AC AAB75062;  
XX  
DT 20-JUL-2001 (first entry)  
XX Human alpha 1-6 fucosyltransferase protein 343-575 SEQ ID NO:4.  
DE Human; alpha 1-6 fucosyltransferase; alpha 1-6 fucT; antibody;  
KW alpha 1-6 fucose transferase; anti-human alpha 1-6 fucT; immunoassay.  
KW  
XX Homo sapiens.  
OS  
XX JP2001011097-A.  
XX  
PD 16-JAN-2001.  
XX  
XX 29-JUN-1999; 99JP-0183569.  
PF  
XX 29-JUN-1999; 99JP-0183569.  
PR  
XX (FURE ) FUJIREBIO KK.  
PA  
XX WPI; 2001-275926/29.  
DR N-PSDB; AAF87953.  
DR  
XX Novel anti-human alpha1-6 fucose transferase antibody useful for  
PT immunoassay -  
XX

Example 2; Page 9-10; lipp; Japanese.  
XX  
XX The present invention describes an anti-human alpha 1-6  
CC fucosyltransferase (alpha 1-6 fucT, also called alpha 1-6 fucose  
CC transferase) antibody. Also described are: (1) a hybridoma producing  
CC the above monoclonal antibody; (2) an immunoassay for detecting human  
CC alpha 1-6 fucT by using the above antibody or its antibody fragment;  
CC and (3) a reagent used for the above immunoassay. The anti-human  
CC alpha 1-6 fucT antibody can be used for immunoassay. The present  
CC sequence represents the human alpha 1-6 fucT protein of residues 343  
CC to 575, which is used in an example from the present invention.  
XX  
SQ Sequence 233 AA;

Query Match 39.1%; Score 1206; DB 22; Length 233;  
Best Local Similarity 96.1%; Pred. No. 4.1e-104;  
Matches 224; Conservative 6; Mismatches 3; Indels 0; Gaps 0;  
QY 343 EKEIEEATKGLGFKHPVIGVHVRRTDKVGTAAFPHEVMVHVEEHQLLARRMQVDKK 402  
DB 1 EKEIEEATKGLGFKHPVIGVHVRRTDKVGTAAFPHEVMVHVEEHQLLARRMQVDKK 60  
QY 403 RVLATDDPTLLKEATKYKSYNTEFTSDNSISWSAGLHNRYSLESLGVLIDHFLSQADF 462.  
DB 61 RVLATDDPTLLKEATKYKSYNTEFTSDNSISWSAGLHNRYSLESLGVLIDHFLSQADF 120  
QY 463 LVCTESSQVCRVAYEIMOTLHPDASANFHSLLDDIYFQGNAHNOIAVPHKPTREEIP 522  
DB 121 LVCTESSQVCRVAYEIMOTLHPDASANFHSLLDDIYFQGNAHNOIAVPHKPTREEIP 180  
QY 523 MEPGDIIGVAGNHWGDSKGINRKLKGTGLYPSYKVRKIEYKTYPTYPEAEK 575  
DB 181 MEPGDIIGVAGNHWGDSKGINRKLKGTGLYPSYKVRKIEYKTYPTYPEAEK 233

RESULT 10  
ABG48804  
ID ABG48804 standard; Peptide; 82 AA.  
XX  
AC ABG48804;  
XX  
DT 25-FEB-2003 (first entry)  
XX Human liver peptide, sEQ ID No 27452.  
DE Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
KW hypercholesterolaemia; coronary heart disease.  
KW  
XX Homo sapiens.  
OS  
XX WO200157273-A2.  
FN  
XX 09-AUG-2001.  
PD  
XX 30-JAN-2001; 2001WO-US00664.  
PF  
XX 04-FEB-2000; 2000US-0380312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0334687.  
PR 27-SEP-2000; 2000US-0336359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX

(MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
FI WPI; 2001-488898/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analysing gene expression in human adult liver -  
XX  
PS Claim 27; SEQ ID No 27452; 658pp; English.  
XX

The invention relates to a single exon nucleic acid probe (SENP) (I) for  
measuring human gene expression in a sample derived from human adult  
liver comprising one of 13109 defined nucleotide sequences given in the  
specification (or complements/ fragments). The probe hybridises at high  
stringency to a nucleic acid molecule expressed in the human adult  
liver. (I) may be used for predicting, measuring and displaying gene  
expression in samples derived from human adult liver. The genes  
identified may be involved in genetic liver diseases such as cirrhosis,  
hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which  
is associated with coronary heart disease. ABG47348-ABG59930 represent  
human liver single exon encoded peptides of the invention.  
Note: The sequence information for this patent does not appear in the  
printed specification but was obtained in electronic format directly  
from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Query Match 14.0%; Score 432; DB 22; Length 82;  
Best Local Similarity 96.3%; Pred. No. 2e-32;  
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 280 EVNDKNQVVELFVDSLHPRPPYPLAVPBDLADRLLRVHGDPVWVWVQFKYLIRPQ 339  
DB 1 EVNDKNQVVELFVDSLHPRPPYPLAVPBDLADRLLRVHGDPVWVWVQFKYLIRPQ 60  
QY 340 PWLEKEIEEATKGLGFKHPVIG 361  
DB 61 PWLEKEIEEATKGLGFKHPVIG 82

RESULT 11

```
ABE28797
ID ABB28797 standard; Peptide; 82 AA.
XX ABB28797;
AC
DT 01-FEB-2002 (first entry)
XX
XX Peptide #1448 encoded by breast cell single exon nucleic acid probe.
DE
XX Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer.
XX
XX Homo sapiens.
OS
XX WO200157271-A2.
PN
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00662.
PF
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
XX Claim 27; SEQ ID NO 11765; 327pp + sequence listing; English.
PS
XX
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 82 AA;
SQ
Query Match 14.0%; Score 432; DB 22; Length 82;
Best Local Similarity 96.3%; Pred. No. 2e-32;
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 280 EVNDKNIQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 339
Db 1 EVKDKNVQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 60
QY 340 PWLEKEIEEATKGLGFKHPVIG 361
Db 61 PWLEKEIEEATKGLGFKHPVIG 82
RESULT 13
ABBI9422
ID ABB19422 standard; Protein; 82 AA.
XX
XX ABB19422;
XX
XX 23-JAN-2002 (first entry)
DT
```

```
ABE28797
ID ABB28797 standard; Peptide; 82 AA.
XX ABB28797;
AC
DT 01-FEB-2002 (first entry)
XX
XX Peptide #1448 encoded by breast cell single exon nucleic acid probe.
DE
XX Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer.
XX
XX Homo sapiens.
OS
XX WO200157271-A2.
PN
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00662.
PF
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
XX Claim 27; SEQ ID NO 11765; 327pp + sequence listing; English.
PS
XX
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 82 AA;
SQ
Query Match 14.0%; Score 432; DB 22; Length 82;
Best Local Similarity 96.3%; Pred. No. 2e-32;
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 280 EVNDKNIQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 339
Db 1 EVKDKNVQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 60
QY 340 PWLEKEIEEATKGLGFKHPVIG 361
Db 61 PWLEKEIEEATKGLGFKHPVIG 82
RESULT 13
ABBI9422
ID ABB19422 standard; Protein; 82 AA.
XX
XX ABB19422;
XX
XX 23-JAN-2002 (first entry)
DT
```

```
ABE28797
ID ABB28797 standard; Peptide; 82 AA.
XX ABB28797;
AC
DT 01-FEB-2002 (first entry)
XX
XX Peptide #1448 encoded by breast cell single exon nucleic acid probe.
DE
XX Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer.
XX
XX Homo sapiens.
OS
XX WO200157271-A2.
PN
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00662.
PF
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
XX Claim 27; SEQ ID NO 11765; 327pp + sequence listing; English.
PS
XX
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 82 AA;
SQ
Query Match 14.0%; Score 432; DB 22; Length 82;
Best Local Similarity 96.3%; Pred. No. 2e-32;
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 280 EVNDKNIQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 339
Db 1 EVKDKNVQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 60
QY 340 PWLEKEIEEATKGLGFKHPVIG 361
Db 61 PWLEKEIEEATKGLGFKHPVIG 82
RESULT 13
ABBI9422
ID ABB19422 standard; Protein; 82 AA.
XX
XX ABB19422;
XX
XX 23-JAN-2002 (first entry)
DT
```

```
ABE28797
ID ABB28797 standard; Peptide; 82 AA.
XX ABB28797;
AC
DT 01-FEB-2002 (first entry)
XX
XX Peptide #1448 encoded by breast cell single exon nucleic acid probe.
DE
XX Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer.
XX
XX Homo sapiens.
OS
XX WO200157271-A2.
PN
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00662.
PF
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
XX Claim 27; SEQ ID NO 11765; 327pp + sequence listing; English.
PS
XX
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 82 AA;
SQ
Query Match 14.0%; Score 432; DB 22; Length 82;
Best Local Similarity 96.3%; Pred. No. 2e-32;
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 280 EVNDKNIQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 339
Db 1 EVKDKNVQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 60
QY 340 PWLEKEIEEATKGLGFKHPVIG 361
Db 61 PWLEKEIEEATKGLGFKHPVIG 82
RESULT 13
ABBI9422
ID ABB19422 standard; Protein; 82 AA.
XX
XX ABB19422;
XX
XX 23-JAN-2002 (first entry)
DT
```



XX DE Protein #1421 encoded by probe for measuring heart cell gene expression.  
XX KW Human; gene expression; heart; microarray; vascular system;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease.  
XX OS Homo sapiens.  
XX PN WO200157274-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US00666.  
XX PR 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0632366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488899/53.  
XX DR Single exon nucleic acid probes for analyzing gene expression in human hearts -  
XX PT hearts -  
XX PS Claim 15; SEQ ID No 21192; 530pp; English.  
XX CC The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart (see ABA21535-ABA41305). The present sequence is a protein encoded by one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease.  
XX CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 82 AA;  
Query Match 14.0%; Score 432; DB 22; Length 82;  
Best Local Similarity 96.3%; Pred. No. 2e-32;  
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 280 EVNDKNIOVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 339  
DB 1 EVKDKNVQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 60  
QY 340 PWLEKEIEEATKGLGFKHPVIG 361  
DB 61 PWLEKEIEEATKGLGFKHPVIG 82  
RESULT 14  
AAM54747  
ID AAM54747 standard; Protein; 82 AA.  
XX AC AAM54747;  
XX OS Homo sapiens.  
XX DT 05-NOV-2001 (first entry)  
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 26852.  
XX KW Human; brain expressed exon; gene expression analysis; probe;

KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.  
XX KW Homo sapiens.  
XX OS Homo sapiens.  
XX PN WO200157275-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US00667.  
XX PR 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0508408.  
XX PR 03-AUG-2000; 2000US-0632366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483446/52.  
XX DR Single exon nucleic acid probes for analyzing gene expression in human brains -  
XX PT brains -  
XX PS Example 4; SEQ ID NO: 26852; 650pp + Sequence Listing; English.  
XX CC The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is a protein encoded by one of the probes of the invention.  
XX SQ Sequence 82 AA;  
Query Match 14.0%; Score 432; DB 22; Length 82;  
Best Local Similarity 96.3%; Pred. No. 2e-32;  
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 280 EVNDKNIOVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 339  
DB 1 EVKDKNVQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 60  
QY 340 PWLEKEIEEATKGLGFKHPVIG 361  
DB 61 PWLEKEIEEATKGLGFKHPVIG 82  
RESULT 15  
AAM67139  
ID AAM67139 standard; Protein; 82 AA.  
XX AC AAM67139;  
XX OS Homo sapiens.  
XX DT 06-NOV-2001 (first entry)  
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 27445.  
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma.  
XX OS Homo sapiens.  
XX PN WO200157276-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US00668.

```
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0623266.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 27445; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention.
XX
SQ Sequence 82 AA;
XX
Query Match 14.0%; Score 432; DB 22; Length 82;
Best Local Similarity 96.3%; Pred. No. 2e-32;
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 280 EYNDKNIQVVELPIVDSLHPRPPYLPVLPEDLADRLRVHGDPAVWVVSQFVKYLIRPQ 339
Db 1 EVKDKNVQVVELPIVDSLHPRPPYLPVLPEDLADRLRVHGDPAVWVVSQFVKYLIRPQ 60
QY 340 PWLEKEIEEATKKLGFGKHPVIG 361
Db 61 PWLEKEIEEATKKLGFGKHPVIG 82
```

Search completed: February 2, 2004, 08:41:47  
Job time : 39.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 08:39:43 ; Search time 15.5 Seconds  
(without alignments)  
1569.596 Million cell updates/sec

Title: US-09-971-773-24  
Perfect score: 3085  
Sequence: 1 MRATGSGWRWIMLILFAWG.....YKREKIETVKYTYPEAEK 575

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3003	97.3	575	3	US-08-913-805A-10
2	3003	97.3	575	3	US-09-442-629-10
3	2910	94.3	575	3	US-08-913-805A-2
4	2910	94.3	575	3	US-09-442-629-2
5	131	4.2	25	3	US-08-913-805A-12
6	131	4.2	25	3	US-09-442-629-12
7	114	3.7	214	5	PCT-US96-08950-2
8	114	3.7	214	5	PCT-US96-09127-2
9	114	3.7	229	3	US-08-630-915A-221
10	113.5	3.7	976	3	US-09-104-324B-4
11	113	3.7	26	3	US-08-913-805A-3
12	113	3.7	26	3	US-09-442-629-3
13	113	3.7	3248	1	US-08-353-700-1
14	113	3.7	3248	5	PCT-US95-16216-1
15	110.5	3.6	644	4	US-09-198-452A-63
16	109	3.5	352	4	US-08-630-915A-12
17	109	3.5	1230	2	US-08-968-542C-35
18	107	3.5	418	4	US-09-198-452A-1056
19	104	3.4	251	4	US-08-630-915A-8
20	104	3.4	677	3	US-08-836-567-2
21	104	3.4	677	4	US-09-606-304-2
22	104	3.4	1197	3	US-08-836-567-12
23	104	3.4	1197	4	US-09-606-304-12
24	103	3.3	575	4	US-09-107-532A-3910
25	102.5	3.3	2482	1	US-08-328-254-6
26	101	3.3	1027	4	US-09-914-259-27
27	101	3.3	1939	4	US-09-310-187A-1

28	101	3.3	2185	2	US-08-822-445-2	Sequence 2, Appli
29	101	3.3	2185	4	US-09-396-540-2	Sequence 2, Appli
30	100	3.2	381	4	US-09-107-532A-3877	Sequence 3877, Ap
31	100	3.2	3878	4	US-09-914-259-11	Sequence 11, Appli
32	99.5	3.2	640	4	US-09-873-404-2	Sequence 2, Appli
33	99.5	3.2	959	4	US-09-914-259-67	Sequence 67, Appli
34	99	3.2	366	4	US-09-750-580-3	Sequence 3, Appli
35	99	3.2	366	4	US-09-599-360B-89	Sequence 89, Appli
36	99	3.2	1037	4	US-09-134-001C-4794	Sequence 4794, Ap
37	98	3.2	1005	2	US-08-935-450-2	Sequence 2, Appli
38	98	3.2	1588	5	PCT-US93-07261-11	Sequence 11, Appli
39	98	3.2	1663	5	PCT-US93-07261-16	Sequence 16, Appli
40	97.5	3.2	353	1	US-07-752-101A-35	Sequence 35, Appli
41	97.5	3.2	466	4	US-09-610-401-3	Sequence 3, Appli
42	97.5	3.2	466	4	US-09-167-206-12	Sequence 12, Appli
43	97.5	3.2	1646	4	US-09-535-008-67	Sequence 67, Appli
44	97.5	3.2	1647	4	US-09-535-008-2	Sequence 2, Appli
45	97.5	3.2	1649	4	US-09-535-008-75	Sequence 75, Appli

## ALIGNMENTS

RESULT 1  
US-08-913-805A-10  
; Sequence 10, Application US/08913805A  
; Patent No. 6054304  
; GENERAL INFORMATION:  
; APPLICANT: TANIGUCHI, Naoyuki  
; APPLICANT: UOZUMI, Neofumi  
; APPLICANT: SHIBA, Tetsuo  
; APPLICANT: YANAGIDANI, Shusaku  
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3+ Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
; SOFTWARE: WordPerfect 6.1 Windows  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/913.805A  
; FILING DATE: 7 JAN 1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP97/00171  
; FILING DATE: 23 JAN 1997  
; APPLICATION NUMBER: JP 192260  
; FILING DATE: 22 JUL 1996  
; APPLICATION NUMBER: JP 162813  
; FILING DATE: 24 JUN 1996  
; APPLICATION NUMBER: JP 161648  
; FILING DATE: 21 JUN 1996  
; APPLICATION NUMBER: JP 10365  
; FILING DATE: 24 JAN 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toffenetti, Judith L.  
; REGISTRATION NUMBER: 39,048  
; REFERENCE/DOCKET NUMBER: 2356/3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-429-1776  
; TELEFAX: 202-429-0796  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 575 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear

```

; MOLECULE TYPE: peptide
US-08-913-805A-10

Query Match
Best Local Similarity 97.3%; Score 3003; DB 3; Length 575;
Matches 555; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 MRWTGSRWIMLILFAWGTLIFYGHLVRNDHPDHSSRELSKILAKLERLKQONEDL 60
DB 1 MRPTGSRWIMLILFAWGTLIFYGHLVRNDHPDHSSRELSKILAKLERLKQONEDL 60

QY 61 RMAESLRIPGPIDQGTATGRVRLVEEQVLVAKQIENYKQARNGLGKDHILRRRIE 120
DB 61 RMAESLRIPGPIDQGTATGRVRLVEEQVLVAKQIENYKQARNGLGKDHILRRRIE 120

QY 121 NGAKELWFFLQSELKKLHLEGNELQRADEFLDLGHHERSIMTDLVYLSQTDGAGDWR 180
DB 121 NGAKELWFFLQSELKKLHLEGNELQRADEFLDLGHHERSIMTDLVYLSQTDGAGDWR 180

QY 181 EKEADLTTELVRITYLQNPDKCSKARKLVNINKGCGYGCQLHHVVYCFMIAVGTORT 240
DB 181 EKEADLTTELVRITYLQNPDKCSKARKLVNINKGCGYGCQLHHVVYCFMIAVGTORT 240

QY 241 LILESQNRVATGWTVPVSETCTDRSGLSTGHWSGEVNDKNIQVVELPIVDSLHPR 300
DB 241 LILESQNRVATGWTVPVSETCTDRSGLSTGHWSGEVNDKNIQVVELPIVDSLHPR 300

QY 241 LILESQNRVATGWTVPVSETCTDRSGLSTGHWSGEVNDKNIQVVELPIVDSLHPR 300
DB 241 LILESQNRVATGWTVPVSETCTDRSGLSTGHWSGEVNDKNIQVVELPIVDSLHPR 300

QY 301 PPYLPLAVPEDLADRLRVHGDPAVWVVSQVKYLIRPQPMLEKEIEATKLGFKHPVI 360
DB 301 PPYLPLAVPEDLADRLRVHGDPAVWVVSQVKYLIRPQPMLEKEIEATKLGFKHPVI 360

QY 361 GVHVRTDKVGTAAAFHPIEEYMWVVEHFQLLARRMQVKKRVYLATDDPTLLKEATK 420
DB 361 GVHVRTDKVGTAAAFHPIEEYMWVVEHFQLLARRMQVKKRVYLATDDPTLLKEATK 420

QY 421 YSNYEFISDINSISWSAGLHNRNTENSLRGVILDIHFLSQADPLVCTFSSQVCRVAYEIMQ 480
DB 421 YSNYEFISDINSISWSAGLHNRNTENSLRGVILDIHFLSQADPLVCTFSSQVCRVAYEIMQ 480

QY 481 TLHPDASANFHSLLDIYFGQNAHNOIAVPHKPTREBEIPMEPGDIIGVAGNHWGYS 540
DB 481 TLHPDASANFHSLLDIYFGQNAHNOIAVPHKPTREBEIPMEPGDIIGVAGNHWGYS 540

QY 541 KGINRKLGTGLYPSYKVKREKIEVTKYPTYPEAEK 575
DB 541 KGINRKLGTGLYPSYKVKREKIEVTKYPTYPEAEK 575

RESULT 2
US-09-442-629-10
; Sequence 10, Application US/09442629
; Patent No. 6291219
; GENERAL INFORMATION:
; APPLICANT: TANIGUCHI, Naoyuki
; UOZUMI, Naofumi
; SHIBA, Tetsuo
; YANAGIDANI, Shusaku
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESS: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: US
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3+ Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
; SOFTWARE: WordPerfect 6.1 Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/442,629
;
; FILING DATE: 18-No. 6291219-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/913.805A
; FILING DATE: 7 JAN 1998
; APPLICATION NUMBER: PCT/JP97/00171
; FILING DATE: 23 JAN 1997
; APPLICATION NUMBER: JP 192260
; FILING DATE: 22 JUL 1996
; APPLICATION NUMBER: JP 162813
; FILING DATE: 24 JUN 1996
; APPLICATION NUMBER: JP 161648
; FILING DATE: 21 JUN 1996
; APPLICATION NUMBER: JP 10365
; FILING DATE: 24 JAN 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Toffenetti, Judith L.
; REGISTRATION NUMBER: 39,048
; REFERENCE/DOCKET NUMBER: 2356/3
; TELEPHONE: 202-429-1776
; TELEFAX: 202-429-0796
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 575 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-442-629-10

Query Match 97.3%; Score 3003; DB 3; Length 575;
Best Local Similarity 96.5%; Pred. No. 4.5e-274;
Matches 555; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 MRWTGSRWIMLILFAWGTLIFYGHLVRNDHPDHSSRELSKILAKLERLKQONEDL 60
DB 1 MRPTGSRWIMLILFAWGTLIFYGHLVRNDHPDHSSRELSKILAKLERLKQONEDL 60

QY 61 RMAESLRIPGPIDQGTATGRVRLVEEQVLVAKQIENYKQARNGLGKDHILRRRIE 120
DB 61 RMAESLRIPGPIDQGTATGRVRLVEEQVLVAKQIENYKQARNGLGKDHILRRRIE 120

QY 121 NGAKELWFFLQSELKKLHLEGNELQRADEFLDLGHHERSIMTDLVYLSQTDGAGDWR 180
DB 121 NGAKELWFFLQSELKKLHLEGNELQRADEFLDLGHHERSIMTDLVYLSQTDGAGDWR 180

QY 181 EKEADLTTELVRITYLQNPDKCSKARKLVNINKGCGYGCQLHHVVYCFMIAVGTORT 240
DB 181 EKEADLTTELVRITYLQNPDKCSKARKLVNINKGCGYGCQLHHVVYCFMIAVGTORT 240

QY 241 LILESQNRVATGWTVPVSETCTDRSGLSTGHWSGEVNDKNIQVVELPIVDSLHPR 300
DB 241 LILESQNRVATGWTVPVSETCTDRSGLSTGHWSGEVNDKNIQVVELPIVDSLHPR 300

QY 241 LILESQNRVATGWTVPVSETCTDRSGLSTGHWSGEVNDKNIQVVELPIVDSLHPR 300
DB 241 LILESQNRVATGWTVPVSETCTDRSGLSTGHWSGEVNDKNIQVVELPIVDSLHPR 300

QY 301 PPYLPLAVPEDLADRLRVHGDPAVWVVSQVKYLIRPQPMLEKEIEATKLGFKHPVI 360
DB 301 PPYLPLAVPEDLADRLRVHGDPAVWVVSQVKYLIRPQPMLEKEIEATKLGFKHPVI 360

QY 361 GVHVRTDKVGTAAAFHPIEEYMWVVEHFQLLARRMQVKKRVYLATDDPTLLKEATK 420
DB 361 GVHVRTDKVGTAAAFHPIEEYMWVVEHFQLLARRMQVKKRVYLATDDPTLLKEATK 420

QY 421 YSNYEFISDINSISWSAGLHNRNTENSLRGVILDIHFLSQADPLVCTFSSQVCRVAYEIMQ 480
DB 421 YSNYEFISDINSISWSAGLHNRNTENSLRGVILDIHFLSQADPLVCTFSSQVCRVAYEIMQ 480

QY 481 TLHPDASANFHSLLDIYFGQNAHNOIAVPHKPTREBEIPMEPGDIIGVAGNHWGYS 540
DB 481 TLHPDASANFHSLLDIYFGQNAHNOIAVPHKPTREBEIPMEPGDIIGVAGNHWGYS 540

QY 541 KGINRKLGTGLYPSYKVKREKIEVTKYPTYPEAEK 575
DB 541 KGINRKLGTGLYPSYKVKREKIEVTKYPTYPEAEK 575
```

## RESULT 3

US-08-913-805A-2  
; Sequence 2, Application US/08913805A  
; Patent No. 6054304  
; GENERAL INFORMATION:  
; APPLICANT: TANIGUCHI, Naoyuki  
; APPLICANT: UOZUMI, Naofumi  
; APPLICANT: SHIBA, Tetsuo  
; APPLICANT: YANAGIDANI, Shusaku  
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3+ Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
; SOFTWARE: WordPerfect 6.1 Windows  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/913,805A  
; FILING DATE: 7 JAN 1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP97/00171  
; FILING DATE: 23 JAN 1997  
; APPLICATION NUMBER: JP 192260  
; FILING DATE: 22 JUL 1996  
; APPLICATION NUMBER: JP 162813  
; FILING DATE: 24 JUN 1996  
; APPLICATION NUMBER: JP 161648  
; FILING DATE: 21 JUN 1996  
; APPLICATION NUMBER: JP 10365  
; FILING DATE: 24 JAN 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toffenetti, Judith L.  
; REGISTRATION NUMBER: 39,048  
; REFERENCE/DOCKET NUMBER: 2356/3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-429-1776  
; TELEFAX: 202-429-0796  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 575 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-913-805A-2

Query Match 94.3%; Score 2910; DB 3; Length 575;

Best Local Similarity 93.6%; Pred. No. 2.6e-265; Mismatches 23; Indels 0; Gaps 0;

	Matches	538;	Conservative	14;
QY	1	MRAWTGSRWIMLILFANGTLLFYIGGHLVRDNDHSDHSSRELSKILAKLERLKQONEDL	60	
Db	1	MRPWTGSRWIMLILFANGTLLFYIGGHLVRDNDHSDHSSRELSKILAKLERLKQONEDL	60	
QY	61	RMASLSRIPEGPIDQGTATGRVRLVEQLVAKQIQENYKQANGLGKQKHILRRRIE	120	
Db	61	RMASLSRIPEGPIDQGPASGRVRLVEQLVAKQIQENYKQANGLGKQKHILRRRIE	120	
QY	121	NCAKELWFLQSELAKKLHLEGNELORHADELILLDGHHSRISIMTDLYVLSQTDGADWR	180	
Db	121	NCAKELWFLQSELAKKLHLEGNELORHADELILLDGHHSRISIMTDLYVLSQTDGADWR	180	
QY	181	EKEAKDLTELVRRTITYLQNPDKCSKAKKLVNKNKGGCGYCOLHHVVVCFMAYGTORT	240	

	Db	181	EKEAKDLTELVRRTITYLQNPDKCSKAKKLVNKNKGGCGYCOLHHVVVCFMAYGTORT <th>240</th>	240
QY	241	LILESQNRVATGMEVFRPVSETCTDRSGLSTGHSWGEVNDKNIQVVELPIVDSLHPR	300	
Db	241	LALSHNRVATGMEVFRPVSETCTDRSGLSTGHSWGEVNDKNIQVVELPIVDSLHPR	300	
QY	301	PPYLPVAVPEDLADRLLRVHGDPAVWVVSQFVKYLIRPQWLEKEIEBATKLGFKHPVI	360	
Db	301	PPYLPVAVPEDLADRLVRVHGDPAVWVVSQFVKYLIRPQWLEKEIEBATKLGFKHPVI	360	
QY	361	GVHVRRTDKVCTBAAPHPIEYMWVHEHFOILLARMQVDKRVYLATDDPTLLKEATK	420	
Db	361	GVHVRRTDKVGAFAFPIEYTVHVEEDFOLLARMQVDKRVYLATDDPTLLKEATK	420	
QY	421	YSNYEFISDSISWSAGLHNYTENSRLGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ	480	
Db	421	YPSYEFISDSISWSAGLHNYTENSRLGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ	480	
QY	481	TLHPDASANFSLDDIYVFGGONAHNQIAVYPHPRTEEEIPMBPGDIIGVAGNHWGYS	540	
Db	481	ALHPDASANFSLDDIYVFGGONAHNQIAVYPHPRTEEEIPMBPGDIIGVAGNHWGYP	540	
QY	541	KGINRKLGTGLYPSYKVKREKIEYKYPTYPEAEK	575	
Db	541	KGVNRKLGTGLYPSYKVKREKIEYKYPTYPEADK	575	

## RESULT 4

US-09-442-629-2  
; Sequence 2, Application US/09442629  
; Patent No. 6291219  
; GENERAL INFORMATION:  
; APPLICANT: TANIGUCHI, Naoyuki  
; APPLICANT: UOZUMI, Naofumi  
; APPLICANT: SHIBA, Tetsuo  
; APPLICANT: YANAGIDANI, Shusaku  
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3+ Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
; SOFTWARE: WordPerfect 6.1 Windows  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/442,629  
; FILING DATE: 18-NO. 6291219-1999  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/913,805A  
; FILING DATE: 7 JAN 1998  
; APPLICATION NUMBER: PCT/JP97/00171  
; FILING DATE: 23 JAN 1997  
; APPLICATION NUMBER: JP 192260  
; FILING DATE: 22 JUL 1996  
; APPLICATION NUMBER: JP 162813  
; FILING DATE: 24 JUN 1996  
; APPLICATION NUMBER: JP 161648  
; FILING DATE: 21 JUN 1996  
; APPLICATION NUMBER: JP 10365  
; FILING DATE: 24 JAN 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toffenetti, Judith L.  
; REGISTRATION NUMBER: 39,048  
; REFERENCE/DOCKET NUMBER: 2356/3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-429-1776  
; TELEFAX: 202-429-0796

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 575 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-442-629-2

Query Match 94.3%; Score 2910; DB 3; Length 575;  
Best Local Similarity 93.6%; Pred. No. 2.6e-265;  
Matches 538; Conservative 14; Mismatches 23; Indels 0; Gaps 0;  
QY 1 MRWGTGSRWIMLLPFWGTLFYIGGHLVRNDHPDHSRSLSKILAKLERLKQONEDL 60  
DB 1 MRWGTGSRWIMLLPFWGTLFYIGGHLVRNDHSDHSSRSLSKILAKLERLKQONEDL 60  
QY 61 RMAESLRIPGPIDGTATGRVRLVEQLVKAQGIENYKQARNGLGKQHEILRRRIE 120  
DB 61 RMAESLRIPGPIDQGPASGRVRLAEQFMKAQGIENYKQTKNGPKGKDEILRRRIE 120  
QY 121 NGAKELWFFLQSELKKLHLEGNELQORHADELTLDLGHERSIMTDLVYLSOTDAGDWR 180  
DB 121 NGAKELWFFLQSELKKLHLEGNELQORHADELTLDLGHERSIMTDLVYLSOTDAGDWR 180  
QY 181 EKEAKDLTELVRRTYTLQNPDKCSKARKLVNINKGCGYGCQLHHVYVCFMAYGTQRT 240  
DB 181 EKEAKDLTELVRRTYTLQNPDKCSKARKLVNINKGCGYGCQLHHVYVCFMAYGTQRT 240  
QY 241 LILESQWRYATGWTETPRPVSCTDRSGISTGHWSGEVNDKNIQVPELPIVDSLHPR 300  
DB 241 LALESHNRYATGWTETPRPVSCTDRSGISTGHWSGEVNDKNIQVPELPIVDSVHPR 300  
QY 301 PPVPLAVPEDLADRLRLVHGDPVWVWSQFVKYLIRPOPWLEKEIEEATKLGFKHPVI 360  
DB 301 PPVPLAVPEDLADRLRLVHGDPVWVWSQFVKYLIRPOPWLEKEIEEATKLGFKHPVI 360  
QY 361 GVHVRTDKVGTAAFPHTPEYMHVVEHFQLLARRMQVDKRVYLATDDPTLLKEATK 420  
DB 361 GVHVRTDKVGAFAFPHTPEYTVHVEEDFQLLARRMQVDKRVYLATDDPALLKEATK 420  
QY 421 YSNYEFISNSISWSAGLHNRYTENSRLRGVILDIHFLSQADFLVCTFSQVCRVAYEIMQ 480  
DB 421 YPSYEFISNSISWSAGLHNRYTENSRLRGVILDIHFLSQADFLVCTFSQVCRVAYEIMQ 480  
QY 481 TLHPDASANFSLDDIYFEGGNAHNOIAVYPHKRTEIEIPMEPGDIIIGVAGNWDGYS 540  
DB 481 ALHPDASANFRSLDDIYFEGGNAHNOIAVYPHQRTGEIPMEPGDIIIGVAGNWDGYP 540  
QY 541 KGINRKLGTGLYPSYKVRKIEYKPYTPYPAEK 575  
DB 541 KGVNRKLGTGLYPSYKVRKIEYKPYTPYPAEK 575

## RESULT 5

US-08-913-805A-12  
; Sequence 12, Application US/08913805A  
; Patent No. 6054304  
; GENERAL INFORMATION:  
; APPLICANT: TANIGUCHI, Naoyuki  
; APPLICANT: UOZUMI, Naofumi  
; APPLICANT: SHIBA, Tetsuo  
; APPLICANT: YANAGIDANI, Shusaku  
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20036  
; COMPUTER READABLE FORM:

MEDIUM TYPE: 3+ Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/913,805A  
FILING DATE: 7 JAN 1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP97/00171  
FILING DATE: 23 JAN 1997  
APPLICATION NUMBER: JP 192260  
FILING DATE: 22 JUL 1996  
APPLICATION NUMBER: JP 162813  
FILING DATE: 24 JUN 1996  
APPLICATION NUMBER: JP 161648  
FILING DATE: 21 JUN 1996  
APPLICATION NUMBER: JP 10365  
FILING DATE: 24 JAN 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Toffenetti, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 2356/3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-913-805A-12

Query Match 4.2%; Score 131; DB 3; Length 25;  
Best Local Similarity 100.0%; Pred. No. 1e-05;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 KLGFKHPVIGVHVVRTDKVGTAAAF 376  
DB 1 KLGFKHPVIGVHVVRTDKVGTAAAF 25

## RESULT 6

US-09-442-629-12  
; Sequence 12, Application US/09442629  
; Patent No. 6291219  
; GENERAL INFORMATION:  
; APPLICANT: TANIGUCHI, Naoyuki  
; APPLICANT: UOZUMI, Naofumi  
; APPLICANT: SHIBA, Tetsuo  
; APPLICANT: YANAGIDANI, Shusaku  
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3+ Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
; SOFTWARE: WordPerfect 6.1 Windows  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/442,629  
; FILING DATE: 18-NO. 6291219-1999  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/913,805A  
; FILING DATE: 7 JAN 1998  
; APPLICATION NUMBER: PCT/JP97/00171

FILING DATE: 23 JAN 1997  
APPLICATION NUMBER: JP 192260  
FILING DATE: 22 JUL 1996  
APPLICATION NUMBER: JP 162813  
FILING DATE: 24 JUN 1996  
APPLICATION NUMBER: JP 161648  
FILING DATE: 21 JUN 1996  
APPLICATION NUMBER: JP 10365  
FILING DATE: 24 JAN 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Toffenetti, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 2356/3  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 12:  
US-09-442-629-12

Query Match 4.2%; Score 131; DB 3; Length 25;  
Best Local Similarity 100.0%; Pred. No. 1e-05;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 KLGFKHPVIGVHVRTDKVGTEAAF 376  
Db 1 KLGFKHPVIGVHVRTDKVGTEAAF 25

RESULT 7  
PCT-US96-08950-2  
Sequence 2, Application PC/TUS9608950  
GENERAL INFORMATION:  
APPLICANT: OSTEOSA INC.  
TITLE OF INVENTION: Osteoclast Growth Regulatory Factor  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/08950  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Paul C. Steinhart  
REGISTRATION NUMBER: 30,806  
REFERENCE/DOCKET NUMBER: FP-ND 2121  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 214 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US96-08950-2

Query Match 3.7%; Score 114; DB 5; Length 214;

Best Local Similarity 38.0%; Pred. No. 0.012;  
Matches 27; Conservative 12; Mismatches 22; Indels 10; Gaps 2;  
Qy 509 AVYPHKPRTTEEIPMEPGDIIGVAG----NHWGYSKGINKRLKGTGLYPSYKVKREK1ET 564  
Db 19 ALYTFEPTDDELYFEEGDIIYITDMSDTNWKTSK-----GRTGLIPSNYYVAEQAES 72  
Qy 565 VKYPTYPEAK 575  
Db 73 IDNPLHEAAKR 83

RESULT 8  
PCT-US96-09127-2  
Sequence 2, Application PC/TUS9609127  
GENERAL INFORMATION:  
APPLICANT: OSTEOSA INC.  
TITLE OF INVENTION: Osteoclast Growth Regulatory Factor  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/09127  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Paul C. Steinhart  
REGISTRATION NUMBER: 30,806  
REFERENCE/DOCKET NUMBER: FP-ND 2122  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 214 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US96-09127-2

Query Match 3.7%; Score 114; DB 5; Length 214;  
Best Local Similarity 38.0%; Pred. No. 0.012;  
Matches 27; Conservative 12; Mismatches 22; Indels 10; Gaps 2;  
Qy 509 AVYPHKPRTTEEIPMEPGDIIGVAG----NHWGYSKGINKRLKGTGLYPSYKVKREK1ET 564  
Db 19 ALYTFEPTDDELYFEEGDIIYITDMSDTNWKTSK-----GRTGLIPSNYYVAEQAES 72  
Qy 565 VKYPTYPEAK 575  
Db 73 IDNPLHEAAKR 83

RESULT 9  
US-08-630-915A-221  
Sequence 221, Application US/08630915A  
Patent No. 6309820  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: HOFFMAN, No. 6309820h  
APPLICANT: KAY, Brian K.  
APPLICANT: FOWLKES, Dana M.  
APPLICANT: MCCONNELL, Stephen J.





ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3+ Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/913,805A  
FILING DATE: 7 JAN 1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP97/00171  
FILING DATE: 23 JAN 1997  
APPLICATION NUMBER: JP 192260  
FILING DATE: 22 JUL 1996  
APPLICATION NUMBER: JP 162813  
FILING DATE: 24 JUN 1996  
APPLICATION NUMBER: JP 161648  
FILING DATE: 21 JUN 1996  
APPLICATION NUMBER: JP 10365  
ATTORNEY/AGENT INFORMATION:  
NAME: Toffenetti, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 2356/3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-913-805A-3

Query Match 3.7%; Score 113; DB 3; Length 26;  
Best Local Similarity 88.0%; Pred. No. 0.00054;  
Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
QY 102 KOARNGLGKDHILRRRIENGAKEL 126  
DB 1 KOTKNGPGKHILRRRIENGAKEL 25  
||:|||||||||||||||||||  
||:|||||||||||||||||||  
  
RESULT 12  
US-09-442-629-3  
Sequence 3, Application US/09442629  
Patent No. 6291219  
GENERAL INFORMATION:  
APPLICANT: TANIGUCHI, Naoyuki  
UOZUMI, Naofumi  
SHIBA, Tetsuo  
YANAGIDANI, Shusaku  
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenyon & Kenyon  
STREET: 1025 Connecticut Avenue, N.W., Suite 600  
CITY: Washington  
STATE: DC  
COUNTRY: US  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3+ Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/442,629  
FILING DATE: 18-NO. 6291219-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/913,805A

FILING DATE: 7 JAN 1998  
APPLICATION NUMBER: PCT/JP97/00171  
FILING DATE: 23 JAN 1997  
APPLICATION NUMBER: JP 192260  
FILING DATE: 22 JUL 1996  
APPLICATION NUMBER: JP 162813  
FILING DATE: 24 JUN 1996  
APPLICATION NUMBER: JP 161648  
FILING DATE: 21 JUN 1996  
APPLICATION NUMBER: JP 10365  
FILING DATE: 24 JAN 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Toffenetti, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 2356/3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-442-629-3  
  
Query Match 3.7%; Score 113; DB 3; Length 26;  
Best Local Similarity 88.0%; Pred. No. 0.00054;  
Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
QY 102 KOARNGLGKDHILRRRIENGAKEL 126  
DB 1 KOTKNGPGKHILRRRIENGAKEL 25  
||:|||||||||||||||||||  
||:|||||||||||||||||||  
  
RESULT 13  
US-08-353-700-1  
Sequence 1, Application US/08353700  
Patent No. 5599919  
GENERAL INFORMATION:  
APPLICANT: YEN, TIMOTHY J.  
APPLICANT: RATTNER, JEROME B.  
TITLE OF INVENTION: NUCLEIC ACID ENCODING A  
TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,  
TITLE OF INVENTION: AND METHODS OF USE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN  
STREET: 1601 MARKET STREET, SUITE 720  
CITY: PHILADELPHIA  
STATE: PA  
COUNTRY: USA  
ZIP: 19103-2307  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/353,700  
FILING DATE: 09-DEC-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: REED, JANET E.  
REGISTRATION NUMBER: 36,252  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 563-4100  
TELEFAX: (215) 563-4044  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3248 amino acids  
TYPE: amino acid

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
;
US-08-353-700-1

Query Match          3.7%; Score 113; DB 1; Length 3248;
Best Local Similarity 19.3%; Pred. No. 1;
Matches 102; Conservative 86; Mismatches 176; Indels 164; Gaps 26;

QY 29 LVRDNDHPDHSRSLKILAKLERLKKQONEDL---RRMAESLRIPGIDQGTATGRVRV 85
DB 343 LVRTTAQYDQAS---TKYTALEQKLKLTEDLSQORQNAESARCS----- 384

QY 86 LEEQL-VKAKEQIENYKKQARNGLGKDHILRRRIE-----NGAKELWFFLOSELKCLKH 139
DB 385 LEQIKKEKEFEQELSQRQSFOTLQECIQMKARLTQELQQAQNMHNVLQAEILDKLT 444

QY 140 LEGNELORHADIILLDLGHHERSIMTDLYYLSQTDGADWREKEAKDLTELVRITVLO 199
DB 445 VK-QOLENNLEEFKQKLCRAEAFQ-----ASQIKENELRRSMEEMKKNLLK 492

QY 200 NPKDCSKARKLVNINCGCGYCOLHHVYCFMIAYGRTILLESQNW-----RYATGGM 255
DB 493 SHSE-QKARE-VCHLE-----AELKNIKQC-----LNQSQNFABEMKAKNTSQ 533

QY 256 ETVPFVSECTDRSGLSTGHSWGEVNDKNIQVVELPIV-----DSLHPRPPYLP 305
DB 534 ETMLRDLQEKINQOE-----NSLTLEKLVADLEKQDCSDLLKKEHHI- 581

QY 306 LAVPEDLADRLLRVHGDPVAV-----WVVSQFVKYLIRPQWL 342
DB 582 -----EQNDKLSKTEKESKALLSLELKKEEYELKEEKTLPSCWSENEKLLTQ-----M 633

QY 343 EKEIEEATKGLGFKHPVIGVHVRRTDKVGTAAFPPIEYVMHVE---EHFOLLARRMQ- 398
DB 634 ESEKENLQSKINHLETCL-----KTQIKS-----HEYNERVRLTMDRENLSVEIRNLHN 684

QY 399 -VDKRRVYLATDDPTLLK-EAKTKSYNEFISDINSISWAGLHNYTEN-----SLRGV 450
DB 685 VLDSKSVEVETQKLAYMELOQKAEFSDQK-----HKEIENMCLKTSQLTGQ 731

QY 451 ILDIHFLSQADFLVCTFSSQVCRVAYEIM--QTLHPDASANFHSLLDDI 496
DB 732 VEDLEHKLQL-----LSNEIMDKRCYQDLHAEYESLRDL 766
```

```
RESULT 14
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Rattner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochore Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
```

```
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US95-16216-1

Query Match          3.7%; Score 113; DB 5; Length 3248;
Best Local Similarity 19.3%; Pred. No. 1;
Matches 102; Conservative 86; Mismatches 176; Indels 164; Gaps 26;

QY 29 LVRDNDHPDHSRSLKILAKLERLKKQONEDL---RRMAESLRIPGIDQGTATGRVRV 85
DB 343 LVRTTAQYDQAS---TKYTALEQKLKLTEDLSQORQNAESARCS----- 384

QY 86 LEEQL-VKAKEQIENYKKQARNGLGKDHILRRRIE-----NGAKELWFFLOSELKCLKH 139
DB 385 LEQIKKEKEFEQELSQRQSFOTLQECIQMKARLTQELQQAQNMHNVLQAEILDKLT 444

QY 140 LEGNELORHADIILLDLGHHERSIMTDLYYLSQTDGADWREKEAKDLTELVRITVLO 199
DB 445 VK-QOLENNLEEFKQKLCRAEAFQ-----ASQIKENELRRSMEEMKKNLLK 492

QY 200 NPKDCSKARKLVNINCGCGYCOLHHVYCFMIAYGRTILLESQNW-----RYATGGM 255
DB 493 SHSE-QKARE-VCHLE-----AELKNIKQC-----LNQSQNFABEMKAKNTSQ 533

QY 256 ETVPFVSECTDRSGLSTGHSWGEVNDKNIQVVELPIV-----DSLHPRPPYLP 305
DB 534 ETMLRDLQEKINQOE-----NSLTLEKLVADLEKQDCSDLLKKEHHI- 581

QY 306 LAVPEDLADRLLRVHGDPVAV-----WVVSQFVKYLIRPQWL 342
DB 582 -----EQNDKLSKTEKESKALLSLELKKEEYELKEEKTLPSCWSENEKLLTQ-----M 633

QY 343 EKEIEEATKGLGFKHPVIGVHVRRTDKVGTAAFPPIEYVMHVE---EHFOLLARRMQ- 398
DB 634 ESEKENLQSKINHLETCL-----KTQIKS-----HEYNERVRLTMDRENLSVEIRNLHN 684

QY 399 -VDKRRVYLATDDPTLLK-EAKTKSYNEFISDINSISWAGLHNYTEN-----SLRGV 450
DB 685 VLDSKSVEVETQKLAYMELOQKAEFSDQK-----HKEIENMCLKTSQLTGQ 731

QY 451 ILDIHFLSQADFLVCTFSSQVCRVAYEIM--QTLHPDASANFHSLLDDI 496
DB 732 VEDLEHKLQL-----LSNEIMDKRCYQDLHAEYESLRDL 766
```

```
RESULT 15
US-09-198-452A-63
; Sequence 63, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
```

;; CURRENT APPLICATION NUMBER: US/09/198.452A  
;; CURRENT FILING DATE: 1998-11-24  
;; NUMBER OF SEQ ID NOS: 6849  
;; SEQ ID NO 63  
;; LENGTH: 644  
;; TYPE: PRT  
;; ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-63

Query Match 3.6%; Score 110.5; DB 4; Length 644;  
Best Local Similarity 21.0%; Pred. No. 0.14; Indels 125; Gaps 20;  
Matches 103; Conservative 72; Mismatches 191; Indels 125; Gaps 20;  
Qy 43 LSKILAKLERLQNMEDLRRMAESLRIPEGPIDQCTA-----TGRVRLERQLVKAKEQI 97  
Db 75 ISGFLLLLERREVSGLGEGPTG--IPVPSAEPSSSEIQKQKAKQILDLQLPQELDQL 132  
Qy 98 ENYKKQARNGLG--KDHEILRRRIENGAK-----LWFFLQSELKKL-----KHL 140  
Db 133 DTDIQHVLSCILGKLDLCKDRGLLKDAKEKLQVDFVWKDMMMEFVELQQVMDQESRYL 192  
Qy 141 EG--NELQRHADEILLDLGHERSMTDLYLSQTDGAGDWREKEAKDLTELQVRRITYL 198  
Db 193 EGLIHEVQSIHAKLFVDDVNTIRSHLGESGYPSPSEDVRGBELLKRFPAK---EVVAR---FM 246  
Qy 199 QNPDKCSKARKLVNCNKGCGYCOLHHVYCFMIAYGTQRTLILESQNMRYA----- 251  
Db 247 KVTED---IRKIAPFNKN--AYGAKN---AFDKAFGLSETCLYKSLTKSYRDTFCDYK 298  
Qy 252 -----TGGWETVPRPVSETCTDRSGLSTGHWSGEVNDKNIQVVLPFVDSLHPR- 300  
Db 299 RAKILPDENNARSARAEQRFREVKDHWEDLN--ETVFWVKE--DGRIDIEVLTAVGWGPDY 354  
Qy 301 PPYLPPLAVPED-----LADRLLRVHGDPVWVWSQFVKYLIRPQWLEKEIEATK--- 351  
Db 355 PEHLILEKRRKDKVMSHQLWEATMRVKEAEVTYSVAR-VAFEKDGSGQQNKKFQEKTKERL 413  
Qy 352 -----KLGFKHPVIGVHVRTD---KVGTEAAPHPPIEE----- 381  
Db 414 RCLKDLRDOECHRQERLEKLTALYPEVSVSVVETERERKFNLEKAYGNLEERYQSVVQD 473  
Qy 382 -----YMHVVEEHFQLLARMQVDRKRVYLATDDPTLLKEAK 418  
Db 474 QEDYWTEQKNREAEFRAGTKVRSMEVEAEHLQILENLEDCYKELSKAETFALGVVEREA 533  
Qy 419 TKYSNYEFISD 429  
Db 534 TEEIEYTLISD 544

Search completed: February 2, 2004, 08:44:37  
Job time : 16.5 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 08:40:23 ; Search time 217 Seconds  
(without alignments)  
550.686 Million cell updates/sec

Title: US-09-971-773-24  
Perfect score: 3085  
Sequence: 1 MRAWTGSWRIMLILFWANGT.....YKVRKTIETVKYPTYPEAEK 575

Scoring table: BL0SUM62  
Gapop 10.0 , Gapext 0.5

Searched: 789580 seqs, 207824079 residues

Total number of hits satisfying chosen parameters: 789580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3085	100.0	575	11	US-09-971-773-24
2	3024	98.0	575	11	US-09-971-773-23
3	2992	97.0	575	9	US-09-839-136-10
4	2903	94.1	575	9	US-09-839-136-2
5	2658	86.2	515	15	US-10-106-698-4658
6	432	14.0	82	9	US-09-864-761-34720
7	426	13.8	81	9	US-09-864-761-46107
8	127	4.1	1055	12	US-10-369-493-12504
9	123	4.0	485	12	US-10-104-047-3419
10	122	4.0	25	9	US-09-839-136-12
11	118.5	3.8	976	12	US-10-117-937-596
12	116.5	3.8	919	12	US-10-369-493-22825
13	116.5	3.8	1992	12	US-10-369-493-6527
14	116	3.8	678	12	US-10-369-493-5427
15	116	3.8	1285	15	US-10-205-823-273

16	116	3.8	1285	15	US-10-177-293-317	Sequence 317, App
17	115	3.7	1137	12	US-10-369-493-6931	Sequence 6931, Ap
18	115	3.7	1156	12	US-10-369-493-43	Sequence 43, Appl
19	114.5	3.7	888	11	US-09-893-519A-73	Sequence 73, Appl
20	114	3.7	223	9	US-09-879-957-221	Sequence 221, App
21	113.5	3.7	2020	12	US-10-369-493-5128	Sequence 5128, Ap
22	113.5	3.7	2020	12	US-10-369-493-5129	Sequence 5129, Ap
23	113	3.7	26	9	US-09-839-136-3	Sequence 3, Appli
24	113	3.7	405	12	US-10-094-749-3191	Sequence 3191, Ap
25	113	3.7	1959	12	US-10-028-248A-36	Sequence 36, Appl
26	111.5	3.6	1961	12	US-10-028-248A-105	Sequence 105, App
27	111	3.6	1441	12	US-10-412-897-3	Sequence 3, Appli
28	111	3.6	1597	14	US-10-017-216-6	Sequence 6, Appli
29	111	3.6	1641	14	US-10-017-216-5	Sequence 5, Appli
30	111	3.6	2055	14	US-10-017-216-4	Sequence 4, Appli
31	110.5	3.6	644	12	US-10-289-762-63	Sequence 63, Appl
32	110.5	3.6	1979	15	US-10-205-823-419	Sequence 419, App
33	110	3.6	586	12	US-10-097-111-286	Sequence 286, App
34	109.5	3.5	262	12	US-10-032-189-79	Sequence 79, Appl
35	109.5	3.5	1179	12	US-10-369-493-1330	Sequence 1330, Ap
36	109.5	3.5	1179	12	US-10-369-493-20394	Sequence 20394, A
37	109.5	3.5	1978	12	US-10-094-466-64	Sequence 64, Appl
38	109	3.5	352	9	US-09-879-957-12	Sequence 12, Appl
39	109	3.5	1001	12	US-10-262-445-66	Sequence 66, Appl
40	109	3.5	1068	12	US-10-262-445-64	Sequence 64, Appl
41	108.5	3.5	521	12	US-10-369-493-5956	Sequence 5956, Ap
42	108	3.5	880	12	US-10-369-493-21643	Sequence 21643, A
43	107	3.5	418	12	US-10-289-762-1056	Sequence 1056, Ap
44	107	3.5	863	12	US-10-369-493-10297	Sequence 10297, A
45	106	3.4	807	12	US-10-132-350-42	Sequence 42, Appl

ALIGNMENTS

RESULT 1

US-09-971-773-24  
; Sequence 24, Application US/09971773  
; Publication No. US20030115614A1  
; GENERAL INFORMATION:  
; APPLICANT: Yutaka KANDA  
; APPLICANT: Mitsuo SATOH  
; APPLICANT: Kazuyasu NAKAMURA  
; APPLICANT: Kazuhisa UCHIDA  
; APPLICANT: Toyohide SHINKAWA  
; APPLICANT: Naoko YAMANE  
; APPLICANT: Motoo YAWASAKI  
; APPLICANT: NO. US20030115614A1uo HANAI  
; TITLE OF INVENTION: ANTIBODY COMPOSITION-PRODUCING CELL  
; FILE REFERENCE: 249-202  
; CURRENT APPLICATION NUMBER: US/09/971,773  
; PRIOR FILING DATE: 2002-08-30  
; PRIOR APPLICATION NUMBER: JP 2000-308526  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 60/268,926  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 24  
; LENGTH: 575  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; US-09-971-773-24

Query Match	100.0%;	Score 3085;	DB 11;	Length 575;
Best Local Similarity	100.0%;	Pred. No. 2.3e-271;		
Matches	575;	Conservative	0;	Mismatches 0;
Indels	0;	Gaps	0;	
Qy	1	MRAWTGSWRIMLILFWANGTLLFYIGGHLVRDNDHPHSSRELSKIILAKLERLKOQNEDL	60	
Db	1	MRAWTGSWRIMLILFWANGTLLFYIGGHLVRDNDHPHSSRELSKIILAKLERLKOQNEDL	60	
Qy	61	RMMAESLRIPGPDQGTATGRVRLVEQLVKAKEQIENYKQARNGLGKDKHILRRRIE	120	

Db 61 RMAESLRIPGPIDQGTATGRVRLVEQLVKAKEQIENYKQARNGLGKDHETLRRRIE 120  
Qy 121 NGAKELWFFLOSELKHLKLEGNELQRADEILLDLGHHERSIMTDLVYLSQTDGAGDWR 180  
Db 121 NGAKELWFFLOSELKHLKLEGNELQRADEILLDLGHHERSIMTDLVYLSQTDGAGDWR 180  
Qy 181 EKEAKDLTELVRRTITLQNPDKCSKARKLVNKNKGGYGCQLHHVYVCFMIAVGTQRT 240  
Db 181 EKEAKDLTELVRRTITLQNPDKCSKARKLVNKNKGGYGCQLHHVYVCFMIAVGTQRT 240  
Qy 241 LILESQNRWYATGCHETVFRPVSETCTDRSGLSGTGHSGEVNDKNIQVVELPIVDSLHPR 300  
Db 241 LILESQNRWYATGCHETVFRPVSETCTDRSGLSGTGHSGEVNDKNIQVVELPIVDSLHPR 300  
Qy 301 PYPPLAVPEDLADRLRLRHVGDPAVWVSQFVKYLIRPQWLEKEIEEATKKGFKHPVI 360  
Db 301 PYPPLAVPEDLADRLRLRHVGDPAVWVSQFVKYLIRPQWLEKEIEEATKKGFKHPVI 360  
Qy 361 GVHVRTDKVGTAAAFHPIEEYVMVHVEHFQLLARRMQVKKRVYLATDDPTLLKEATK 420  
Db 361 GVHVRTDKVGTAAAFHPIEEYVMVHVEHFQLLARRMQVKKRVYLATDDPTLLKEATK 420  
Qy 421 YSNYEFISDNTSWSAGLHNRYTENSRLGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
Db 421 YSNYEFISDNTSWSAGLHNRYTENSRLGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
Qy 481 TLHPDASANFHSLLDDIYFVGQNAHQIAVYPHKPRTREEIPMEPGDIIGVAGNHWGYS 540  
Db 481 TLHPDASANFHSLLDDIYFVGQNAHQIAVYPHKPRTREEIPMEPGDIIGVAGNHWGYS 540  
Qy 541 KGINRKLKGTGLYPSYKVKREKIEVTKYPTYPEAEK 575  
Db 541 KGINRKLKGTGLYPSYKVKREKIEVTKYPTYPEAEK 575

## RESULT 2

US-09-971-773-23  
; Sequence 23, Application US/09971773  
; Publication No. US20030115614A1  
; GENERAL INFORMATION:  
; APPLICANT: Yutaka KANDA  
; APPLICANT: Mitsuo SATOH  
; APPLICANT: Kazuyasu NAKAMURA  
; APPLICANT: Kazuhisa UCHIDA  
; APPLICANT: Toyohide SHINKAWA  
; APPLICANT: Naoko YAMANE  
; APPLICANT: Motoo YASAKI  
; APPLICANT: No. US20030115614A1uo HANAI  
; TITLE OF INVENTION: ANTIBODY COMPOSITION-PRODUCING CELL  
; FILE REFERENCE: 249-202  
; CURRENT APPLICATION NUMBER: US/09/971,773  
; CURRENT FILING DATE: 2002-08-30  
; PRIOR APPLICATION NUMBER: JP 2000-308526  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 60/268,926  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 23  
; LENGTH: 575  
; TYPE: PRT  
; ORGANISM: Cricetus griseus  
US-09-971-773-23

Query Match 98.0%; Score 3024; DB 11; Length 575;  
Best Local Similarity 97.6%; Pred. No. 8.1e-266;  
Matches 561; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MRATGSRWIMLILFANGTLLFYIGGHLVRDNDHPDHSRELSKILAKLERLKOQNEDL 60  
Db 1 MRATGSRWIMLILFANGTLLFYIGGHLVRDNDHPDHSRELSKILAKLERLKOQNEDL 60

Qy 61 RMAESLRIPGPIDQGTATGRVRLVEQLVKAKEQIENYKQARNGLGKDHETLRRRIE 120  
Db 61 RMAESLRIPGPIDQGTATGRVRLVEQLVKAKEQIENYKQARNGLGKDHETLRRRIE 120  
Qy 121 NGAKELWFFLOSELKHLKLEGNELQRADEILLDLGHHERSIMTDLVYLSQTDGAGDWR 180  
Db 121 NGAKELWFFLOSELKHLKLEGNELQRADEILLDLGHHERSIMTDLVYLSQTDGAGDWR 180  
Qy 181 EKEAKDLTELVRRTITLQNPDKCSKARKLVNKNKGGYGCQLHHVYVCFMIAVGTQRT 240  
Db 181 EKEAKDLTELVRRTITLQNPDKCSKARKLVNKNKGGYGCQLHHVYVCFMIAVGTQRT 240  
Qy 241 LILESQNRWYATGCHETVFRPVSETCTDRSGLSGTGHSGEVNDKNIQVVELPIVDSLHPR 300  
Db 241 LILESQNRWYATGCHETVFRPVSETCTDRSGLSGTGHSGEVNDKNIQVVELPIVDSLHPR 300  
Qy 301 PYPPLAVPEDLADRLRLRHVGDPAVWVSQFVKYLIRPQWLEKEIEEATKKGFKHPVI 360  
Db 301 PYPPLAVPEDLADRLRLRHVGDPAVWVSQFVKYLIRPQWLEKEIEEATKKGFKHPVI 360  
Qy 361 GVHVRTDKVGTAAAFHPIEEYVMVHVEHFQLLARRMQVKKRVYLATDDPTLLKEATK 420  
Db 361 GVHVRTDKVGTAAAFHPIEEYVMVHVEHFQLLARRMQVKKRVYLATDDPTLLKEATK 420  
Qy 421 YSNYEFISDNTSWSAGLHNRYTENSRLGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
Db 421 YSNYEFISDNTSWSAGLHNRYTENSRLGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
Qy 481 TLHPDASANFHSLLDDIYFVGQNAHQIAVYPHKPRTREEIPMEPGDIIGVAGNHWGYS 540  
Db 481 TLHPDASANFHSLLDDIYFVGQNAHQIAVYPHKPRTREEIPMEPGDIIGVAGNHWGYS 540  
Qy 541 KGINRKLKGTGLYPSYKVKREKIEVTKYPTYPEAEK 575  
Db 541 KGINRKLKGTGLYPSYKVKREKIEVTKYPTYPEAEK 575

## RESULT 3

US-09-839-136-10  
; Sequence 10, Application US/09839136  
; Patent No. US20020081694A1  
; GENERAL INFORMATION:  
; APPLICANT: Naoyuki TANIGUCHI et al.  
; TITLE OF INVENTION: ALPHA 1-6 FUCOSYLTRANSFERASE  
; FILE REFERENCE: 2356-7  
; CURRENT APPLICATION NUMBER: US/09/839,136  
; CURRENT FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: 09/442,629  
; PRIOR FILING DATE: 1999-11-18  
; PRIOR APPLICATION NUMBER: 08/913,805  
; PRIOR FILING DATE: 1998-01-07  
; PRIOR APPLICATION NUMBER: PCT/JP97/00171  
; PRIOR FILING DATE: 1997-01-23  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 575  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-839-136-10

Query Match 97.0%; Score 2992; DB 9; Length 575;  
Best Local Similarity 96.2%; Pred. No. 6.6e-263;  
Matches 553; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MRATGSRWIMLILFANGTLLFYIGGHLVRDNDHPDHSRELSKILAKLERLKOQNEDL 60  
Db 1 MRATGSRWIMLILFANGTLLFYIGGHLVRDNDHPDHSRELSKILAKLERLKOQNEDL 60

Qy 61 RMAESLRIPGPIDQGTATGRVRLVEQLVKAKEQIENYKQARNGLGKDHETLRRRIE 120  
Db 61 RMAESLRIPGPIDQGTATGRVRLVEQLVKAKEQIENYKQARNGLGKDHETLRRRIE 120

```
Qy 121 NGAKELWFFLOSELKLLKHLGNEQLORHADEILLDLGHHERSIMTDLYLSSQTDGAGDWR 180
Db 121 NGAKELWFFLOSELKLLKHLGNEQLORHADEILLDLGHHERSIMTDLYLSSQTDGAGDWR 180
Qy 181 EKEAKDLTELQVRITTYLQNPDKCSKAKLVNCNKGCGYGCQLHHVVCYCFMIAYGRT 240
Db 181 EKEAKDLTELQVRITTYLQNPDKCSKAKLVNCNKGCGYGCQLHHVVCYCFMIAYGRT 240
Qy 241 LILESQNRWYATGWEVTFRVPVSETCTDRSGLSGTGHSGEVNDKNIQVVELPIVDSLHPR 300
Db 241 LILESQNRWYATGWEVTFRVPVSETCTDRSGLSGTGHSGEVNDKNIQVVELPIVDSLHPR 300
Qy 301 PPYLPLAVPEDLADRLLRVHGDPAVWVVSQFVKYLIRPQWLEKEIEBEATKLGFKHPVI 360
Db 301 PPYLPLAVPEDLADRLLRVHGDPAVWVVSQFVKYLIRPQWLEKEIEBEATKLGFKHPVI 360
Qy 361 GVHVRRTDKVTEAAFPPIEBEYMHVVEHFOLLARRMQVKKRYLATDDPTLLKEATK 420
Db 361 GVHVRRTDKVTEAAFPPIEBEYMHVVEHFOLLARRMQVKKRYLATDDPTLLKEATK 420
Qy 421 YSNYEFISDINSISWSAGLHNRYTENSURGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480
Db 421 YPNTEFISDINSISWSAGLHNRYTENSURGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480
Qy 481 TLHPDASANFHSLLDDIYFGQNAHNOIAVYPHKPRTEEEIIPMEPGDIIIGVAGNHWGYS 540
Db 481 TLHPDASANFHSLLDDIYFGQNAHNOIAVYAHQPRTADEIPMEPGDIIIGVAGNHWGYS 540
Qy 541 KGINRKLGTGLYPSYKVRKEIETVKYPTYPEAK 575
Db 541 KGVNRKLGRGTGLYPSYKVRKEIETVKYPTYPEAK 575
```

## RESULT 4

```
US-09-839-136-2
; Sequence 2, Application US/09839136
; Patent No. US20020081694A1
; GENERAL INFORMATION:
; APPLICANT: Naoyuki TANIGUCHI et al.
; TITLE OF INVENTION: ALPHA 1-6 FUCOSYLTRANSFERASE
; FILE REFERENCE: 2356-7
; CURRENT APPLICATION NUMBER: US/09/839,136
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 09/442,629
; PRIOR FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: 08/913,805
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: PCT/JP97/00171
; PRIOR FILING DATE: 1997-01-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 575
; TYPE: PRT
; ORGANISM: Pig
US-09-839-136-2
```

```
Query Match 94.1%; Score 2903; DB 9; Length 575;
Best Local Similarity 93.4%; Pred. No. 8.1e-255;
Matches 537; Conservative 14; Mismatches 24; Indels 0; Gaps 0;
Qy 1 MRWTGSRWMTLILFAWGTLILFYGGHLVRNDHPDHSSELKILAKLERLKOQNEDL 60
Db 1 MRWTGSRWMTLILFAWGTLILFYGGHLVRNDHPDHSSELKILAKLERLKOQNEDL 60
Qy 61 RMAESLRIPGPDIDQGTATGRVRLERQLVKAKEQIENYKQARNGLGKDHETLRRRIE 120
Db 61 RMAESLRIPGPDIDQGTATGRVRLERQLVKAKEQIENYKQARNGLGKDHETLRRRIE 120
Qy 121 NGAKELWFFLOSELKLLKHLGNEQLORHADEILLDLGHHERSIMTDLYLSSQTDGAGDWR 180
Db 121 NGAKELWFFLOSELKLLKHLGNEQLORHADEILLDLGHHERSIMTDLYLSSQTDGAGDWR 180
```

```
Qy 181 EKEAKDLTELQVRITTYLQNPDKCSKAKLVNCNKGCGYGCQLHHVVCYCFMIAYGRT 240
Db 181 EKEAKDLTELQVRITTYLQNPDKCSKAKLVNCNKGCGYGCQLHHVVCYCFMIAYGRT 240
Qy 241 LILESQNRWYATGWEVTFRVPVSETCTDRSGLSGTGHSGEVNDKNIQVVELPIVDSLHPR 300
Db 241 LILESQNRWYATGWEVTFRVPVSETCTDRSGLSGTGHSGEVNDKNIQVVELPIVDSLHPR 300
Qy 301 PPYLPLAVPEDLADRLLRVHGDPAVWVVSQFVKYLIRPQWLEKEIEBEATKLGFKHPVI 360
Db 301 PPYLPLAVPEDLADRLLRVHGDPAVWVVSQFVKYLIRPQWLEKEIEBEATKLGFKHPVI 360
Qy 361 GVHVRRTDKVTEAAFPPIEBEYMHVVEHFOLLARRMQVKKRYLATDDPTLLKEATK 420
Db 361 GVHVRRTDKVTEAAFPPIEBEYMHVVEHFOLLARRMQVKKRYLATDDPTLLKEATK 420
Qy 421 YSNYEFISDINSISWSAGLHNRYTENSURGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480
Db 421 YPSYEFISDINSISWSAGLHNRYTENSURGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480
Qy 481 TLHPDASANFHSLLDDIYFGQNAHNOIAVYPHKPRTEEEIIPMEPGDIIIGVAGNHWGYS 540
Db 481 ALHPDASANFRSLLDDIYFGQNAHNOIAIYHPQRTGEIIPMEPGDIIIGVAGNHWGYP 540
Qy 541 KGINRKLGTGLYPSYKVRKEIETVKYPTYPEAK 575
Db 541 KGVNRKLGRGTGLYPSYKVRKEIETVKYPTYPEADK 575
```

## RESULT 5

```
US-10-106-698-4658
; Sequence 4658, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4658
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-4658
```

```
Query Match 86.2%; Score 2658; DB 15; Length 515;
Best Local Similarity 96.1%; Pred. No. 1.3e-232;
Matches 491; Conservative 12; Mismatches 8; Indels 0; Gaps 0;
```

```
Qy 65 ESURIPGPDIDQGTATGRVRLERQLVKAKEQIENYKQARNGLGKDHETLRRRIENGAK 124
Db 5 KSLRIPGPDIDQGTATGRVRLERQLVKAKEQIENYKQARNGLGKDHETLRRRIENGAK 64
Qy 125 ELWFFLOSELKLLKHLGNEQLORHADEILLDLGHHERSIMTDLYLSSQTDGAGDWRKEA 184
Db 65 ELWFFLOSELKLLKHLGNEQLORHADEILLDLGHHERSIMTDLYLSSQTDGAGDWRKEA 124
Qy 185 KDLTELQVRITTYLQNPDKCSKAKLVNCNKGCGYGCQLHHVVCYCFMIAYGRTTILE 244
Db 125 KDLTELQVRITTYLQNPDKCSKAKLVNCNKGCGYGCQLHHVVCYCFMIAYGRTTILE 184
Qy 245 SONRYATGWEVTFRVPVSETCTDRSGLSGTGHSGEVNDKNIQVVELPIVDSLHPRPPL 304
Db 185 SONRYATGWEVTFRVPVSETCTDRSGLSGTGHSGEVNDKNIQVVELPIVDSLHPRPPL 244
```

Qy 305 PLAVPEDLADRLRLRVHGDPAVWVWSQFVKYLIRPOPWLEKEIEEATKGLGPKHPVIGHV 364  
Db 245 PLAVPEDLADRLRLRVHGDPAVWVWSQFVKYLIRPOPWLEKEIEEATKGLGPKHPVIGHV 304  
Qy 365 RTIDKVGTEAAHPTEEEYVHVVEEHFQLLARMQVDKRVVLYATDDPTLLKEAKTKYSNY 424  
Db 305 RTIDKVGTEAAHPTEEEYVHVVEEHFQLLARMQVDKRVVLYATDDPTLLKEAKTKYSNY 364  
Qy 425 EFISDNSISWSAGLHNRHTENSRLGVILDIHFLSQADFLVCTFFSQVCRVAYEIMQTLHP 484  
Db 365 EFISDNSISWSAGLHNRHTENSRLGVILDIHFLSQADFLVCTFFSQVCRVAYEIMQTLHP 424  
Qy 485 DASAFHSLDDIYFEGGNAHNAQIAVPHKPTTEBIPMEPGDDIIGVAGNHWGYSKGIN 544  
Db 425 DASAFHSLDDIYFEGGNAHNAQIAVPHKPTTEBIPMEPGDDIIGVAGNHWGYSKGIN 484  
Qy 545 RKLGTGLYPSYKVKREKIEETVQPYTPAEK 575  
Db 485 RKLGTGLYPSYKVKREKIEETVQPYTPAEK 515

RESULT 6  
US-09-864-761-34720  
; Sequence 34720, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecmica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

; SEQ ID NO 34720  
; LENGTH: 82  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AL109847.1  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.2  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6  
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.6  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.1  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.9  
; OTHER INFORMATION: SWISSPROT HIT: Q9VLZ7, EVALUE 6.80e+00  
; OTHER INFORMATION: EST\_HUMAN HIT: AW02060.1, EVALUE 7.00e-44  
US-09-864-761-34720  
Query Match 14.0%; Score 432; DB 9; Length 82;  
Best Local Similarity 96.3%; Pred. No. 1.7e-31;  
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Qy 280 EVNDKNIQVVELPIVDSLHPRPPYPLAVPEDLADRLRLRVHGDPAVWVWSQFVKYLIRPQ 339  
Db 1 EVKDKNQVVELPIVDSLHPRPPYPLAVPEDLADRLRLRVHGDPAVWVWSQFVKYLIRPQ 60  
Qy 340 PWLEKEIEEATKGLGPKHPVIG 361  
Db 61 PWLEKEIEEATKGLGPKHPVIG 82  
RESULT 7  
US-09-864-761-46107  
; Sequence 46107, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecmica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1



PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
SEQ ID NO 46107  
LENGTH: 81  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL109847.3  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3  
OTHER INFORMATION: SWISSPROT HIT: Q9VLZ7, EVALUE 6.40e+00  
OTHER INFORMATION: EST\_HUMAN HIT: AW002060.1, EVALUE 5.00e-43  
US-09-864-761-46107

Query Match 13.8%; Score 426; DB 9; Length 81;  
Best Local Similarity 96.3%; Pred. No. 5.8e-31;  
Matches 78; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 280 EVNDKNIQVVELPIVDSLHPPYPLAVPDLADRLRLRVHGDPAVWVSQFVKYLIRPQ 339  
Db 1 EVKKNVQVVELPIVDSLHPPYPLAVPDLADRLRVHGDPAVWVSQFVKYLIRPQ 60

Qy 340 PWLEKEIEATKGLGFKHPVI 360  
Db 61 PWLEKEIEATKGLGFKHPVI 81

RESULT 8  
US-10-369-493-12504  
Sequence 12504, Application US/10369493  
Publication No. US2003023675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 12504  
LENGTH: 1055  
TYPE: PRT  
ORGANISM: Aspergillus nidulans  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)-(1055)  
OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-12504

Query Match 4.1%; Score 127; DB 12; Length 1055;  
Best Local Similarity 25.6%; Pred. No. 0.041;  
Matches 44; Conservative 38; Mismatches 6; Indels 30; Gaps 7;

Qy 42 ELSKILAKLERLKQNEDLRRMAESLRIPGPDGTATGRVRLVEQLVKAKEQIENYK 101  
Db 771 KIKIMIBIDSAKRSLADARRVQELSAHKP--SKTDSARVKVLEEQIVGLEQIEDLR 828

Qy 102 KQARNGLGKDHKYLRRRI-----ENGAKELWFFLOSEL-----KKLKH 139  
Db 829 SQ-KGGIEETEQALQNKIMEVGVRLSQKAKVDGLKEQISLLSEISNAEVARSKNEKL 887

Qy 140 LEGNELOR-HADHILLDGHHSIMTDLVYLSQTDGAGDWREK--EAKDIT 188  
Db 888 IKCHENARAFAEKELANVAEELQKLNEDV--ANOANDASGWKEKVDQAQDVS 937

RESULT 9  
US-10-104-047-3419  
Sequence 3419, Application US/10104047  
Publication No. US20030236392A1  
GENERAL INFORMATION:  
APPLICANT: HELIX RESEARCH INSTITUTE  
TITLE OF INVENTION: No. US20030236392A1el full length cDNA  
FILE REFERENCE: HI-A0105  
CURRENT APPLICATION NUMBER: US/10/104,047  
CURRENT FILING DATE: 2002-03-25  
PRIOR APPLICATION NUMBER:  
PRIOR FILING DATE:  
NUMBER OF SEQ ID NOS: 4096  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3419  
LENGTH: 485  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-104-047-3419

Query Match 4.0%; Score 123; DB 12; Length 485;  
Best Local Similarity 20.3%; Pred. No. 0.03;  
Matches 92; Conservative 70; Mismatches 140; Indels 152; Gaps 23;

Qy 45 KILAKLERLK-QNEDLRRMAESLRIPGPDGTATGRVRLVEQLVKAKEQIENYKQ 103  
Db 13 KKVTKAERLKLQEEERLKEE-----EASLK-----YSKE 45

Qy 104 ARNGLGKDHKYLRRRIENGAKELWFFLOSELKKLHLEGNELORHADI-----LLDLG-- 157  
Db 46 -----EMERLEIQRIEKRW-----HRLKADLERNEELEYLLERCFF 86

Qy 158 -----HHERSIMTDLVYLSQTDGAGD-----WREKBAKOLTELVR----- 193  
Db 87 EAEKLIKQETKLLSQWKHYIQCDGSPDPSVAQEMTFTSLWKEKTNETFEEVIEKSKVLN 146

Qy 194 -----RITYLQNPXDCSKARKLVNKGCGYGLHHVYCFMIAVGTQRTLLSQNW 248  
Db 147 LIEKLFILLETLP-PCDLQDNIIQYESI---LQLOQLLH---LKFQVATEILLKQ--- 196

Qy 249 RYATGGWETVFRPVSETCDRSGLSTGHSVNDKNIQVVELPIVDSLHPPYPLAV 308  
Db 197 -----ASTLAD---LDSGNMEKVIKDN-----VTLVYWANLKNPRLHRSVRP 236

Qy 309 PE-----DLADRLRLRVHGD-----PAVWV---VSQFVKYLIR-PQP 340  
Db 237 SETQIGFEIPRILATSDIAVRLHHTHYDVSALHPVSTPSPKEYTSVAVTELKVDVKVNEK 296

Qy 341 WLEKIEEATKGLGFKHPVIGVHVRTDKVGTAAFPHEIPEYVHVVEHFOLL--ARMQ 398  
Db 297 AISKEVEESKQ-----QERGSHLIOEEELKVEEEOGDIEVKMSSABEESAICEREMK 351

Qy 399 VDKRVVYLATDDPTLLKKEAKTKYSNYEFISDINSI 432  
Db 352 VLSETVSAA---QLLLVENSE--KPDFFDNVV 380

RESULT 10  
US-09-839-136-12  
Sequence 12, Application US/09839136  
Patent No. US20020081694A1  
GENERAL INFORMATION:  
APPLICANT: Naoyuki TANIGUCHI et al.  
TITLE OF INVENTION: ALFA 1-6 FUCOSYLTRANSFERASE

```
; FILE REFERENCE: 2356-7
; CURRENT APPLICATION NUMBER: US/09/839,136
; CURRENT FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 09/442,629
; PRIOR FILING DATE: 1999-11-18
; PRIOR APPLICATION NUMBER: 08/913,805
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: PCT/JP97/00171
; PRIOR FILING DATE: 1997-01-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Primer
US-09-839-136-12
```

```
Query Match 4.0%; Score 122; DB 9; Length 25;
Best Local Similarity 96.0%; Pred. No. 0.00043;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 352 KLGFKHPVIGVHVRTRDKVGTAAAF 376
Db 1 KLGFKHPVIGVHVRTRDKVGTCAAF 25
```

## RESULT 11

```
US-10-117-937-596
; Sequence 596, Application US/10117937
; Publication No. US20030220239A1
; GENERAL INFORMATION:
; APPLICANT: CTL IMMUNO THERAPIES CORP.
; APPLICANT: SIMARD, John, J.L.
; APPLICANT: DIAMOND, David, C.
; APPLICANT: LIU, Liping
; APPLICANT: XIE, Zhidong
; TITLE OF INVENTION: EPITOPE SEQUENCES
; FILE REFERENCE: CTLIMM.027A
; CURRENT APPLICATION NUMBER: US/10/117,937
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/282,211
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: US 60/337,017
; PRIOR FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: US 60/363,210
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 602
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 596
; LENGTH: 976
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-117-937-596
```

```
Query Match 3.8%; Score 118.5; DB 12; Length 976;
Best Local Similarity 19.5%; Pred. No. 0.22;
Matches 91; Conservative 66; Mismatches 171; Indels 139; Gaps 18;
```

```
QY 45 KILAKLERLUQNEDLRMAESLRIPGIDQGTATGRV-RVLEBQLVKAKEIQENYKKQ 103
Db 157 KVSLEEGIQENKOL-----IKENNATRLCNLLKTKETCARSAEKKYKYE 203
QY 104 -----ARNGLOK---DHEILRRRIENGAKELWFFLOSELKKLKHLEG-----NE 144
Db 204 REETROYMDLANNTEKMTATGEGLRVQAEENSRLEMHFKLKEDYKIOHLEQYKKEIND 263
QY 145 LORHADEILLDGHHSRIMTDLYL--SOTDAGDWRE-----K 182
Db 264 KEQVSLLIQITEKENK-MKDLTFLEESRDKVQLEBTKKQSENKQSTIEKQHHLTK 322
QY 183 EAKDLTELVRRI-TYLNQPKDCSKARKLVNCNINCKGGYGCQLHHVYCFMAYGTQRTL 241
```

```
Db 323 ELEDIKVSLQRSVSTQKALEEDLQIATKTCQLTE-----EKETQ 362
QY 242 ILESONWRYATGCGWETVFRPVSETCTDRSGLSGTGHWSGEVNDKNIQVWELPIVDSLHPRP 301
Db 363 MEESNKARAAHSFVVTEFE--TTVCSLEELLRTQORLEKNEDQLKILTMELQKK----- 415
QY 302 PYLPLAVPEDLADLLRVHGDPAVWWSQFVKYILRPOQWLEKEIEBATKKGKHPVIG 361
Db 416 -----SSELEEMTKLTNNKEVELEELKKVLGSKETLLY 448
QY 362 VHVRTDKVGTAAAFHPLEEMVHVEHF--QLLARRMOVDKRVYL--ATDDPTLLKE 416
Db 449 ENKQ-----FEKIAELKGTQELIGLLOQAREKEVHDLFIQLTAITTSQYYSKE 498
QY 417 AK-----TKYSNYBPISD-NSISWSAGLHNRYTENSRLGVLDI 454
Db 499 VKDLKTELENEKLNKLTSHCNKLS----LENKELTQETSDMTLEL 541
```

## RESULT 12

```
US-10-369-493-22825
; Sequence 22825, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
```

```
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
```

```
; CURRENT APPLICATION NUMBER: US/10/369,493
```

```
; PRIOR FILING DATE: 2003-02-28
```

```
; PRIOR APPLICATION NUMBER: US 60/360,039
```

```
; PRIOR FILING DATE: 2002-02-21
```

```
; NUMBER OF SEQ ID NOS: 47374
```

```
; SEQ ID NO 22825
```

```
; LENGTH: 919
```

```
; TYPE: PRT
```

```
; ORGANISM: Aeropyrum pernix
```

```
US-10-369-493-22825
```

```
Query Match 3.8%; Score 116.5; DB 12; Length 919;
Best Local Similarity 22.5%; Pred. No. 0.3;
Matches 108; Conservative 70; Mismatches 160; Indels 143; Gaps 25;
```

```
QY 50 LERLKOQNEDLRRMAESLRIPGIDQGTATGRVVRVLEEQVLVKAKEIQENYKKQARNGLG 109
Db 522 LSRLNQLGEGREL--GFQTP-----DLAKAFQKLRLRLERLEELKULSENSLEEKVRN-LS 575
QY 110 KDHEILR-----RRI---ENGAKELWFFLOSELKKLKHLEGNELQRHADBIILDL 156
Db 576 REEVALREAKTRALEVLQRIQKEEBAREKULTLSSSEKSLERM-----LVSKAEDLATRL 631
QY 157 GHHSRIMTDLY-----YLSQTD---GAGDWREKAKDL-----TELVRRIYTL 198
Db 632 GITAVRSDDLLEKAREALEGVDKELSAIERREARRLKEEAALKWEAEQVMKRLBEL 691
QY 199 QNPDKCSKARKLVNCNINCKGGYGCQLHHVYCFMAYGTQRTLI-LESQWRYATGGWET 257
Db 692 E-----ABEKKLRKEVSRKSEIARLKEV-----QNTLAELDDDRISR----- 728
QY 258 VFRPVSETCTDRSGLSGTGHWSGEVNDKNIQVWELPIVDSLHPRPPYPLA----VPEDLA 313
Db 729 IDREMGELQTRIREMKSRKASGE-----EALKL-----YLPAAASRRIMEIG 771
QY 314 D-----RLLRVHGDPAVWWSQFVKYILRPOQWLEKEIEBATKKGKHPVIGVHVR----- 365
Db 772 EIAVRRLAV-----LEDMNDILSR--FNLDVAGVEIREKAA 807
QY 366 ---RTDKVGTAAAFHPLEEMVHVEHFQ-----LARRMOVDKRVYLATDDPTLLKEA 417
```

Db 808 REIEVKAIGNGAYRPLE--AVSGGERTVLALSFLVNLNKA VGGKLGFLALDEPTANLDE 865  
Qy 418 KTKYNYEFDISNWSAGLHNRRTENSRLGVLDIHFLSQADF--LVC--TFSSOVCR 473  
Db 866 DRRRLVEIRGISV-----EGLVRLVVVTHHEDVRDYADTICLVTRTQOQSR 914  
Qy 474 V 474  
Db 915 V 915

## RESULT 13

US-10-369-493-6527  
; Sequence 6527, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIORITY FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 6527  
; LENGTH: 1992  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-10-369-493-6527

Query Match 3.8%; Score 116.5; DB 12; Length 1992;  
Best Local Similarity 19.1%; Pred. No. 0.97;  
Matches 89; Conservative 69; Mismatches 148; Indels 161; Gaps 19;  
Qy 42 ELKILAKLERLQONEDLRMAE-----SURIEGIDOGTATGRVRV 85  
Db 954 KLSDTGQLEDMOERNEDLAROKKTKDQELSDTKKHVQDLESLRKABEQKQSRDHNR 1013  
Qy 86 LEEQLVAKAEQIENYKQARNGLGKDHILRRRIENGAKELWFFLOSELKXKLE--GN 143  
Db 1014 LQENANQDEAVAKLNKE-----KKHQ-----EESNRKLNEDLQSEEDKVNHLKTRN 1061  
Qy 144 ELQRHAEILLDLGHHERSIMTDLYLSQTDGAGDWREKEAKDLTFLVQRITTYLQNP 203  
Db 1062 KLEQQMDEL-----EENID-----REKRSRGDIEKAKRV-----EGD 1094  
Qy 204 CSKARKLVNINCGCYGQQLHHVYVCPMIAYGTORTLILESQNRWRYATGWTVPVPS 263  
Db 1095 LKVAQENIDEITK-----QKHV-----R 1112  
Qy 264 ETCTDRSGLSGTHWSEVNDKNIQVWELP-IVDSLHPRPPYPLAVPDEL-ADRLLRVHG 321  
Db 1113 ETTLKRKEEDLHHTWAKLAENNSIIAKLQRLKELTARNAEI-----EELAEARNROKS 1168  
Qy 322 DPAVWVWSQFVKYLRPOPWLEKEETEATKLGKPKHPVIGVHV-----R 365  
Db 1169 D-----RSRSEARELEELTERLEQOQGGATAAQLANKREAEIAKLREK 1214  
Qy 366 RTDKVGTAAFPPIE-----EYVHVVEHFQLLAR-----RMQVDKRVVYLATDDP 411  
Db 1215 EEDSLNHETAISSLKRKHGDSVAELTEQLETLQKLKAKSEAKSKLQRLDEESQHATDSE 1274  
Qy 412 T-----LKEATKYSNTEFISD-----NSISWSAGLHNRV-TENS 446  
Db 1275 VRSQDLBKALKTIEVQYSELQTKADEQSRQLQDFPAALKNRNNNS 1321

## RESULT 14

US-10-369-493-5427  
; Sequence 5427, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIORITY FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 5427  
; LENGTH: 678  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-10-369-493-5427

Query Match 3.8%; Score 116; DB 12; Length 678;  
Best Local Similarity 18.1%; Pred. No. 0.21;  
Matches 86; Conservative 81; Mismatches 185; Indels 122; Gaps 16;  
Qy 38 HSRSLSKILAFRLRLKQONEDLRMAESLRIPGPDQGTATGRVVRVLEBOLVAKKQI 97  
Db 255 HLGDMDSIELEFASKMEQADEMRSLABEV-----FNERRAKEBELQGLEIEV 302  
Qy 98 ENYKQARNGLGKDHILRRRIENGAKELWFFLOSELKXKLE----- 138  
Db 303 EEQKKL-----NEAVTHAMDPMQKEKEDLSEAKLLRERVVMEAKNEDLDDRISK 354  
Qy 139 ---HLEGNELQFHA---DEILLDLGHHERSIMTDLYLSQTDGAGDWREKEAKDLTEL 192  
Db 355 YEIEIRSNPLKXKAIQLETLDTLKKQBEKLMEDMQSALTPEA---WRDKSENKMLNA 411  
Qy 193 RRTYVQNPQ-----DCSKARKLVNINCGCYGQQLHHVYVCPMIAYGTQ-----R 239  
Db 412 DLVVIQKHQTKVQKQISLASELHEYDSQGEAQIMAHHTKYLDLSKSTMLDDTTTENPQ 471  
Qy 240 TLILESQNRWRYATGWTVPVPS-----TCTDRSGLSGTHWSEVNDK----- 284  
Db 472 QIVYQDDIEEISDAVVLILRKISANLKKVNLDEQITDLDERGLTL--QTGNVDELKEMH 529  
Qy 285 -----NIQVWELPFI---VDSLHPRPPYPLAVPDELADRLLRVHGDPVWVWSQFVK 333  
Db 530 VRLQBELISIEFMELALNEEIDNLETEK-----KIDQELAGVGNVDS- 574  
Qy 334 YLIRPQPWLEKIEBEATKLGKPKHPVIGVHVVRRTDK--VGTEAAPHPIEYVHVVEHFQ 391  
Db 575 -----GLRQLEERQKLEDEAPARSHENMQOLEANVASIRNELHSIPGYSQH-----K 622  
Qy 392 LLARRMQVDKRVVYLATDDPDLLEAKTKYSNTEFISDINSISWSAGLHNRVTEN 445  
Db 623 MLRERLEAVEKRTAAKSLDMSLRKTE-----IDYEDIKTESIRLOBEYNNMLLTN 672

## RESULT 15

US-10-205-823-273  
; Sequence 273, Application US/10205823  
; Publication No. US20030308963A1  
; GENERAL INFORMATION:  
; APPLICANT: Schlegel, Robert  
; APPLICANT: Monahan, John E.  
; APPLICANT: Endege, Wilson O.  
; APPLICANT: Gannavarapu, Manjula  
; APPLICANT: Gorbacheva, Bella  
; APPLICANT: Hoersch, Sebastian  
; APPLICANT: Kamatkar, Shubhangi

```
; APPLICANT: Wonsey, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 273
; LENGTH: 1285
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-205-823-273

Query Match          3.8%; Score 116; DB 15; Length 1285;
Best Local Similarity 19.7%; Pred. No. 0.56;
Matches 71; Conservative 51; Mismatches 130; Indels 108; Gaps 15;

QY 39 SSRELSKILAKLERLKQONEDLRRMAESLRIPGEPIDQGTATGRVRLVEEQVLVKAKEQIE 98
DB 905 SSELLSALOKKKQOEAEERLRRIQEME-----KERREREDEKRRRKEBEE 953
QY 99 NYKKQARNGLGKDHEILRRRIENGAKELWFFLOSELKLLKHLEGN-----ELQRHADEIL 153
DB 954 RRMKLEMEARKQOEERKREDEKRIQAEVEAQLARQKEEESQQAQVLEQERRDRELA 1013
QY 154 LDLGHERSITMDL-----YLSQTDGAGD--WREKAKDLTELVRRTIYLLQNPXDCSKA 207
DB 1014 LRIAQSEAEILSDAQAADLALRRNDGTRPMTPEQNAKEMSEFLSRGPAVL-----ATKA 1068
QY 208 RKLVCNINKGCGYGCQLHHVVYCFMIAYGTQRTLILESQNWRYATGGWETVFRPVSETCT 267
DB 1069 -----AAGTKK---YDLSKKYAE-----LRDTINTSC 1093
QY 268 D-----RSGLSGTH-WSGEVNDKNIQ-----VVELPIVDSLHPRPPYLPPLA 307
DB 1094 DIELLAACREFFHRLKVVHAWKSKNKNRNTETEORAPKSVTDYDFAPFLNNSPOONPAA 1153
QY 308 -----VPEDLADRLLV-----HGDPVAV---WVVSQFVKYLIRPOPWLEKETE 347
DB 1154 QIPAROREIEMNRQORFFRIPFIRPADQYKDPQSKKKGWYAHF-----DGPWTARQME 1207
```

Search completed: February 2, 2004, 08:52:04  
Job time : 218 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 08:39:43 ; Search time 17 Seconds  
(without alignments)  
3252.764 Million cell updates/sec

Title: US-09-971-773-24

Perfect score: 3085

Sequence: 1 MRWGTGSRWIMLILFANGT.....YKREKIETVKYPTPEAEK 575

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_76:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3003	97.3	575	1 JC5432	glycoprotein 6-alp
2	982	31.8	818	2 T32154	hypothetical prote
3	126	4.1	946	2 S28061	SCPI protein - rat
4	125.5	4.1	845	2 I48176	synaptonemal compl
5	125	4.1	1025	2 H86250	hypothetical prote
6	118	3.8	1173	2 T43527	sp8 protein - firs
7	118	3.8	1727	2 T50073	myosin-like coiled
8	117	3.8	1744	2 F86161	Fluorid protein -
9	117	3.8	1871	2 D96796	probable heat shoc
10	117	3.8	1955	2 T30934	myosin-like protei
11	116.5	3.8	919	2 B72765	hypothetical prote
12	116.5	3.8	1992	1 S02771	myosin heavy chain
13	116	3.8	678	2 H88187	protein C18H9.8 li
14	115.5	3.7	1427	2 S22695	restin - human
15	115	3.7	1137	2 T19414	hypothetical prote
16	115	3.7	1156	2 B70356	chromosome assembl
17	115	3.7	1325	2 S16129	dynein-associated
18	114	3.7	755	2 T34567	hypothetical prote
19	113.5	3.7	1011	2 T13055	dynamain associated
20	113.5	3.7	1023	2 T33338	hypothetical prote
21	113.5	3.7	2020	2 T21174	hypothetical prote
22	113	3.7	1094	2 T13053	dynamain associated
23	112.5	3.6	832	2 H72278	alpha-mannosidase-
24	112.5	3.6	1270	2 T09194	adaptor protein in
25	112.5	3.6	2007	1 B43402	myosin heavy chain
26	111.5	3.6	1281	2 JC5368	dynactin 1 - mouse
27	111.5	3.6	1392	2 A43336	microtubule-vesicl
28	111.5	3.6	1935	1 A37102	myosin beta heavy
29	111	3.6	993	2 S49461	synaptonemal compl

30	111	3.6	1597	2 S68420	citron - mouse
31	111	3.6	2142	2 T08621	centrosome associa
32	110.5	3.6	542	2 E86496	hypothetical prote
33	110.5	3.6	542	2 F81543	hypothetical prote
34	110.5	3.6	978	2 A70387	conserved hypothet
35	110	3.6	581	2 B87768	protein F54Cl.1 li
36	110	3.6	1057	2 F89892	carbamoyl-phosphat
37	109.5	3.5	542	2 F72127	hypothetical prote
38	109.5	3.5	1179	2 F71190	probable chromosom
39	109.5	3.5	1282	2 JF0120	glycoprotein A - m
40	109.5	3.5	1935	1 S06006	myosin beta heavy
41	109	3.5	589	2 T48905	gamma response 1 p
42	109	3.5	1110	2 T49091	myosin response 1 p
43	109	3.5	1934	2 I48153	myosin heavy chain
44	108.5	3.5	521	2 D88640	protein F55A8.2 li
45	108	3.5	380	2 F75103	conserved hypothet

ALIGNMENTS

RESULT 1

JC5432

glycoprotein 6-alpha-L-fucosyltransferase (EC 2.4.1.68) precursor - human

N:Alternate names: alpha 1-6 FucT; GDP-L-fucose:N-acetyl-beta-D-glucosaminyl:6-alpha-L-f

C:Species: Homo sapiens (man)

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 16-Jun-2000

C:Accession: JC5432; PC4322

R:Yanagidani, S.; Uozumi, N.; Ihara, Y.; Miyoshi, E.; Yamaguchi, N.; Taniguchi, N.

J. Biochem. 121, 626-632, 1997

A:Title: Purification and cDNA cloning of GDP-L-Fuc:N-acetyl-beta-D-glucosaminide:alpha

A:Reference number: JC5433; MUID:97279058; PMID:9133635

A:Accession: JC5432

A:Molecule type: mRNA

A:Residues: 1-575 <YAN1>

A:Cross-references: DDBJ:D89289; NID:G2055306; PIDN:BAAL9764.1; PID:G2055307

A:Accession: PC4322

A:Molecule type: Protein

A:Residues: 68-87/352-376;419-432 <YAN2>

C:Comment: This enzyme catalyzes the transfer of fucose from GDP-fucopyranoside to aspar

C:Genetics:

A:Gene: GDB:FUT8

A:Cross-references: GDB:9786294; OMIM:602589

A:Map position: 14q23-14q23

C:Superfamily: human glycoprotein 6-alpha-L-fucosyltransferase

C:Keywords: glycosyltransferase; hexosyltransferase

F:1-22/Domain: signal sequence #status predicted <Sig>

F:23-575/Product: glycoprotein 6-alpha-L-fucosyltransferase #status predicted <MAT>

Query Match 97.3%; Score 3003; DB 1; Length 575;

Best Local Similarity 96.5%; Pred. No. 1.6e-194; Mismatches 9; Indels 0; Gaps 0;

Matches 555; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY	1	MRWGTGSRWIMLILFANGTLLFYIGGHLVRDNDHPDHSRELSKILAKLERLKQONEDL	60
DB	1	MRPWTGSRWIMLILFANGTLLFYIGGHLVRDNDHPDHSRELSKILAKLERLKQONEDL	60
QY	61	RRMAESLRIPSPIDQGTATGRVRLERQLVKAEQIENYKQARNGLGKDHILRRRIE	120
DB	61	RRMAESLRIPSPIDQGTATGRVRLERQLVKAEQIENYKQARNGLGKDHILRRRIE	120
QY	121	NGAKELWFFLOSELKKLHLEGNELORHAEILLDLGHHSIMTDLVYLSQTDGAGDWR	180
DB	121	NGAKELWFFLOSELKKLHLEGNELORHAEILLDLGHHSIMTDLVYLSQTDGAGDWR	180
QY	181	EKEAKDITELVQRITTYLQNPDKCSKARKLVNKNKGCGYQLHHVVCYFMAIYGTORT	240
DB	181	EKEAKDITELVQRITTYLQNPDKCSKARKLVNKNKGCGYQLHHVVCYFMAIYGTORT	240
QY	241	LILSQNRWRYAFGGWETVFRPVSETCTDRSGLSGTGHSWGEVNDKNIQVLEPIVDSLHPR	300
DB	241	LILSQNRWRYAFGGWETVFRPVSETCTDRSGLSGTGHSWGEVNDKNIQVLEPIVDSLHPR	300

QY 301 PYPPLAVPEDLADRLLRVHGDPVAVVWSQFVKYLIRPQPWLEKIEEATKLGKHPVI 360  
 DB 301 PYPPLAVPEDLADRLLRVHGDPVAVVWSQFVKYLIRPQPWLEKIEEATKLGKHPVI 360  
 QY 361 GVHVRTDKVGTAAFPPIEEYVHVHVFQLLARRMOVDKRVYLATDDPTLLKEATK 420  
 DB 361 GVHVRTDKVGTAAFPPIEEYVHVHVFQLLARRMOVDKRVYLATDDPTLLKEATK 420  
 QY 421 YSNYEFISDNSISWSAGLHNRNTENSRLGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
 DB 421 YSNYEFISDNSISWSAGLHNRNTENSRLGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
 QY 481 TLHPDASANFHSDDIYYFGGQNAHNQIAVYVPHKPTREEEIPMERPGDIIIGVAGNHWGYS 540  
 DB 481 TLHPDASANFHSDDIYYFGGQNAHNQIAVYVPHKPTREEEIPMERPGDIIIGVAGNHWGYS 540  
 QY 541 KGINRKLGTGLYPSYKVBREKIEYVKTYPYPAEK 575  
 DB 541 KGVNRKLGRTGLYPSYKVBREKIEYVKTYPYPAEK 575

RESULT 2  
 T32154  
 hypothetical protein C10F3.6 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 R:Accession: T32154  
 R:Davidson, S.; Wohlmann, P.  
 submitted to the EMBL Data Library, September 1997  
 A:Description: The sequence of C. elegans cosmid C10F3.  
 A:Reference number: Z21127  
 A:Accession: T32154  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-818 <DAV>  
 A:Cross-references: EMBL:AF022968; PIDN:AAB69888.1; GSPDB:GN00023; CBSP:C10F3.6  
 A:Experimental source: strain Bristol N2; clone C10F3  
 C:Genetics:  
 A:Gene: CBSP:C10F3.6  
 A:Map position: 5  
 A:Introns: 30/1; 72/2; 235/2; 258/3; 285/2; 315/3; 457/3; 491/1; 544/3; 556/1; 626/1; 64

Query Match 31.8%; Score 982; DB 2; Length 818;  
 Best Local Similarity 36.8%; Pred. No. 3.4e-58;  
 Matches 211; Conservative 122; Mismatches 191; Indels 50; Gaps 11;

QY 3 ANTGSRWITMLILFAWGTLLFYIGHLVRDNDHPDHSRELSKILAKLERLQOQNEDLRR 62  
 DB 6 AAVGTVMVMTMFLYSQL-----SNNQSGDSIRAWRQTKEADIKLQEQNEEDLS 56  
 QY 63 MAESLRPEGPIDQGTATGRVRLER-----QLVKAKEQIENYKQARNGLGRD 111  
 DB 57 ILEKER-----QERNDQHKIMEQSHQLPNPNPNSLPKPEPVKEIISKSILG-PVQ 108  
 QY 112 HEILRRRIENGAKELWFFIQLSELKKLKHLEGNLQRADEIILLDGHHSRIMTDLYLS 171  
 DB 109 QEVOKRMLDRIEMFYLLHSQTIE-----NSTKILLET--QMISLMGLSAQLE 155  
 QY 172 QTDGADWREKAKDITELVQRITVILONPKDCSKARKLVCNINKCGCGCOLHHVYCF 231  
 DB 156 KLEGSEEEFRKATATQIRFIKSIKLOPNKACSEAKTLVCLNDKCEGFCOLHHVYCA 215  
 QY 232 MTAAGTQRTLLI--ESQNRVATGWTFRPVSCTDRSGLSLSTGHWSGE--VNDKNIQ 287  
 DB 216 ITAFATORMMVLKRDGSSWKYSHGWTVPFKLSKCSFDE--AVGNTEAKFPAPSPAR 272  
 QY 288 VVELPVIDSLRPPPYPLAVPEDLADRLLRVHGDPVAVVWSQFVKYLIRPQPWLEKETE 347  
 DB 273 VVSLGIVDSLITKPTPLPOAVPEOLLESLSLHSHPPAFFVGTFTSYLFRFNSATQEKLD 332  
 QY 348 EATKLGKFP-KHPVIGVHVVRTDKVGTAAFPPIEEYVHVHVFQLLARRM-OVDKRVY 405  
 DB 333 KALKSIPLDKGPVIGLQIRTDKVGTEAAFAHALKEYNTEIWFVKVEKROGKPLERRIF 392

QY 406 LATDDPTLLKAKTKYSNYEFISDNSISWSAGLHNRNTENSRLGVILDIHFLSQADFLVC 465  
 DB 393 IASDDPTVPEAKNDYPNYEYVYSTETIAKTAQLANNRYTDASLUMGVITDIYLLSKVNYLVC 452  
 QY 466 TFSSQVCRVAYEIMQTLHPDASANFHSDDIYYFGGQNAHNQIAVYVPHKPTREEEIPMER 525  
 DB 453 TFSSQVCRVAYEIMQTLHPDASANFHSDDIYYFGGQNAHNQIAVYVPHKPTREEEIPMER 525  
 QY 526 GDIIGVAGNHWGYSKGINRKLGTGLYPSYKVR 559  
 DB 513 GDKVGIAGNHWGYSKGTNRQTYKEGVFPYKIR 546

RESULT 3  
 S28061  
 SCPI protein - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 08-Oct-1999  
 R:Accession: S28061  
 R:Meuwissen, R.L.J.; Offenberger, H.H.; Dietrich, A.J.J.; Rieswijk, A.; van Iersel, M.;  
 EMOO J. 11, 5091-5100, 1992  
 A:Title: A coiled-coil related protein specific for synapsed regions of meiotic prophase  
 A:Reference number: S28061; MUID:93099884; PMID:1464329  
 A:Accession: S28061  
 A:Molecule type: mRNA  
 A:Residues: 1-946 <MEU>  
 A:Cross-references: EMBL:X67805; NID:G57212; PIDN:CAA48006.1; PID:G57213  
 C:Genetics:  
 A:Gene: SCPI  
 C:Keywords: DNA binding

Query Match 4.1%; Score 126; DB 2; Length 946;  
 Best Local Similarity 20.6%; Pred. No. 1.9;  
 Matches 109; Conservative 78; Mismatches 180; Indels 162; Gaps 24;

QY 42 ELSKLAKLERLQOQNEDLRRMAESLRPEPIDQGTATGR-VRVLESLQVKAKEIENY 100  
 DB 104 ENKVSLEKLEEEIOENKDL-----IKENNAIRHWCNLLKETCAKSABTSKY 150  
 QY 101 KKQ-----ARNGLGK--DHEILRRRIENGAKELWFFIQLSELKKLKHLEGNELQR 147  
 DB 151 EYEREETQVYVDLNNNIEKMLAFEEELRVQAEARLEWHFKLKEDEKIQHLE-EYQK 209  
 QY 148 HAD-----EILLDLGHHSRIMTDLYL--SQTGADWREKAKDITELVQRITVILQ 199  
 DB 210 EVNNKENQVSLLIQSTENKMKDITFLLESRDKANQLEBK-----TKLODENLKELN 264  
 QY 200 NPKDCSKARKLVCNINKCGCGCOLHHVYCFMIAYGCTQRTLLBSQNRVATGCGWTVF 259  
 DB 265 EKKDHLTS-----ELEDIKMSQMSMTOKTL---EEDLQIAT---KTIY 303  
 QY 260 RPVS-----TCTDRSGLSLSTGHWSGEVND-----283  
 DB 304 QLTEEKAQMEELNKAETHSLVITELKATCTLEELLRTQORLENNEDQLKLITMELQ 363  
 QY 284 -----KNIQVVELPVIDSLRPPPYPLAVPEDLADRLLRVHG-DPAVWVVSQ 330  
 DB 364 KKSSELEEMTKFNKNEVELEELKTI-----LAEDQKLLDEKKQVEKLABELQKBEQ 415  
 QY 331 FVKYLIRPQPWLEKIEEATKLGKHPVIGVHVVRTDKVGTAAFPPIEEYVHVHVEE-- 388  
 DB 416 ELTFLLOTR---EKEIHDL-----VQVTVTKTSE-----EHLVKQVEEMK 453  
 QY 389 -----HFQOLLARR--MQVDKRVVYLATDDPTLLKAKTKYSNYEFISDNSISWSA 436  
 DB 454 TELEKEKLNIELTANSMDLLENKLLQVQASDMVLELKGHQEDIIINCKQBERMLKQLE 513  
 QY 437 GLHNRNTENSRLGVILDI--HFLSQADFLVCTF--SSQVCR-VAYEIMQ 480  
 DB 514 TLEEK--EMNLRDELESVRKEFIOQGVCKLCKDSEENARSIEYEVLK 560



[illegible]





A:Molecule type: DNA  
A:Residues: 1-919 <KAW>  
A:Cross-references: DDBJ:AF000058; NID:g5103388; PIDN:BAA79020.1; PID:g5103499  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE0110  
C:Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032

Query Match 3.8%; Score 116.5; DB 2; Length 919;  
Best Local Similarity 22.5%; Pred. No. 7.8;  
Matches 108; Conservative 70; Mismatches 160; Indels 143; Gaps 25;

QY 50 LERLQKQEDNLRMAESLIPGPIQGTATGRVRLVLEQLVKAKEIENYKQARNGLG 109  
DB 522 LSKLNQLEGRREL--GFQTP--DLAKAEQKLMLRRLERLEAKLNSLEKVRN-LS 575  
QY 110 KHEILR-----RRI---ENGAKELWFFLOSELKHLKHLGNEQLORHADEILLDL 156  
DB 576 REEVALREAKTRALEVLQRLGKEBEAREKLTLSSESKLRLM-----LVSKAEDLATRL 631  
QY 157 GHHERSIMTDLY-----YLSQTD---GAGDWREKEAKOL-----TELVRRTITVL 198  
DB 632 GITAYRSLDDLEKAREALEGVDKLSAIERLEEARLKEEAALKWEAEQVMKLEEL 691  
QY 199 QNPKQCSKARKLVCNINKGCGYCOLHHVYVCFMAYGTQRTLI-LESQNWRYATGGWET 257  
DB 692 E-----AEBKLRKEVSRKSETEARLKEV-----QNTLAELDDRISR----- 728  
QY 258 VFRPVSETCTDRSGLTGSHWSEVNDKNVQVELPIVDLSLHPRPYPLPLA-----VPEDLA 313  
DB 729 IDRMGELQTRIREMKSRKASGE-----EALK-----YLPAAASRRIMEISG 771  
QY 314 D-----RLLRVHGDPVWVVSQFVKYLIRPQPWLEKEIEBATKLGFKHPVIGVHV- 365  
DB 772 ETAYRRLAV-----LEBMDILSR--FNLDVAGVEIREKAA 807  
QY 366 ---RTDKVGTAEAFHPIEEYVHVVEHFQ-----LARRMQVDKRVYLATDPTLLKEA 417  
DB 808 REIEVKAIGGAYRPLE--AVSGGERTVLALSFLVNLNKAVGKLGFLALDEPTANLDE 865  
QY 418 KTKYSNYEFISNSISWSAGLHNRNTENSLRGVILDIHFLSQADP--LVC--TFSSQVCR 473  
DB 866 DRRRLSVEVLRGISV-----EGLVRQLVVVTHEDVRDYADTICLVTRTQQGSR 914  
QY 474 V 474  
DB 915 V 915

RESULT 12  
S02771  
myosin heavy chain A [similarity] - Caenorhabditis elegans  
N:Contains: myosin ATPase (EC 3.6.4.1)  
C:Species: Caenorhabditis elegans  
C>Date: 31-Dec-1993 #sequence\_revision 19-May-2000 #text\_change 19-Apr-2002  
C:Accession: T23622; S02771  
R:Harris, B.  
submitted to the EMBL Data Library, August 1996  
A:Reference number: Z19773  
A:Accession: T23622  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1992 <WIL>  
A:Cross-references: EMBL:Z78159; PIDN:CAB01576.1; GSPDB:GN00023; CESP:K12F2.1  
A:Experimental source: clone K12F2  
R:Diib, N.J.; Maruyama, I.N.; Krause, M.; Karn, J.  
J. Mol. Biol. 205, 603-613, 1989  
A:Title: Sequence analysis of the complete Caenorhabditis elegans myosin heavy chain gen  
A:Reference number: S02771; MUID:89178677; PMID:2926820  
A:Accession: S02771  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-116,140-1992 <DIB>

A:Cross-references: EMBL:X08067; NID:g6798; PIDN:CRA30856.1; PID:g6799  
C:Genetics:  
A:Gene: myo-3; CESP:K12F2.1  
A:Map position: 5  
A:Introns: 46/1; 192/1; 292/1; 468/2; 1921/3  
C:Superfamily: myosin heavy chain; myosin motor domain homology  
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; muscle co  
F:89-802/Domain: myosin motor domain homology <MMOT>  
F:202-209/Region: nucleotide-binding motif A (P-loop)  
F:690-712/Region: actin binding #status predicted  
F:793-807/Region: actin binding #status predicted  
F:875-1992/Domain: coiled coil #status predicted <COI>  
F:875-1189/Region: S2  
F:1190-1992/Region: light meromyosin  
F:153/Modified site: N6,N6-trimethyllysine (lys) #status predicted  
F:208/Binding site: ATP (lys) #status predicted  
F:730,740/Active site: Cys #status predicted

Query Match 3.8%; Score 116.5; DB 1; Length 1992;  
Best Local Similarity 19.1%; Pred. No. 23;  
Matches 89; Conservative 69; Mismatches 148; Indels 161; Gaps 19;

QY 42 ELSKILAKLERLKQONDLRRMAE-----SLRIPEGPIDOGTATGRVVRV 85  
DB 954 KLSDI TQGLDMQERNEDLARQKKTDOELSDTKKHVDLSELKAEQEKSRDHNIRS 1013  
QY 86 LEEQLVKAKEIENYKQARNGLGHDKHILRRRIENGAKELWFFLOSELKHLKLE--GN 143  
DB 1014 LQDEMANQDEAVAKLKE-----KQHQ-----EESNRKLNEDLQSEEDKVNHEKIRN 1061  
QY 144 ELQRHADILLDLGHHERSIMTDLYLSQTDGADWRKEAKDLTELVRRTITVLQNPKD 203  
DB 1062 KLEQOMDEL-----EENID-----REKRSRGDIEAKRV-----EGD 1094  
QY 204 CSKARKLVCNINKGCGYCOLHHVYVCFMAYGTQRTILLESQNWRYATGGWETVFRPV 263  
DB 1095 LKVAQENIDEITK-----QKHDV----- 1112  
QY 264 ETCTDRSGLTGSHWSEVNDKNVQVELP-IVDSLHPRPYPLPLAVPDL-ADRLLRVHG 321  
DB 1113 ETTLKRKEEDLHHTNAKLANNISIIAKUQLRIKELTARNAE-----ESEELEARNRQKS 1168  
QY 322 DPAVWVVSQFVKYLIRPQPWLEKEIEBATKLGFKHPVIGVHV-----R 365  
DB 1169 D-----RSRSEARELEELTELEQGGGATAQAANKKREAIKLRREK 1214  
QY 366 RTDKVGTAEAFHPIE---EYVHVVEHFQLLAR-----RMQVDKRVYLATDPT 411  
DB 1215 EEDSLNHETAISSLRKRGDSVAELTEQLETQLKAKSEAKSKLQRLDEESQATDSE 1274  
QY 412 T-----LLKEAKTKYSNYEFISD---NSISWSAGLHNRN-TENS 446  
DB 1275 VRSRQDLEKALKTIIEVQYSELQTKADEQSRQLQDFALKNRLNNNS 1321

RESULT 13  
H88187  
protein C18H9.8 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
C:Accession: H88187  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_eleg  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: H88187  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-678 <STO>  
A:Cross-references: GB:chr\_II; PID:g722387; GSPDB:GN00020; CESP:C18H9.8  
C:Genetics:

A:Gene: C18H9.8  
A:Map position: 2

Query Match 3.8%; Score 116; DB 2; Length 678;  
Best Local Similarity 18.1%; Pred. No. 5.6;  
Matches 86; Conservative 81; Mismatches 185; Indels 122; Gaps 16;

QY 38 HSRELSKILAKLERLKQONEDLRMAESLRIPGPDQGTATGRVRVLEBQLVKAQEI 97  
DB 255 HLGMSDTELEASMKQADMRSLAEV-----FNERRAKEEELQGLEIEV 302

QY 98 ENYKQARNGLCKDHEILLRRRIENGAKELWFFLOSELKKLK-----138  
DB 303 ERQKDL-----NEAVTHAMDPMQKKEIDELKSEAKLLRERVVMEAKNEDLDLRISK 354

QY 139 ---HLEGNELORHA---DEILLDLGHHSRIMTDLYILSQTDGAGDREKEAKDLTELQ 192  
DB 355 YEIEIRSNPLKKAQLOQETDLTKKQEKLMEDMQSALTEPA---NRDKWSENMKQLNA 411

QY 193 RRTYLNQPK---DCSKARKLVNINKGCGYQOLHHVVYCFMIAVGTQ-----R 239  
DB 412 DLWVIEKQHTVKDQIQLASEBELHYDSQGEAQIMAHHTKYLLDLSSKTMLDTTTNYPO 471

QY 240 TLILESQWRVYATGWETVFRVSE-----TCTDRSGLSTGHWSGEVNDK-----284  
DB 472 QIVYQODIEEFSADVILRKISANLKKVNLDEQITDLDERGLTL---QTGNVDLEKMH 529

QY 285 -----NIQVVELPI---VDSLHPRPPYLPVAPVEDLADRLLRVHGDPVAVWVVSQFVK 333  
DB 530 VRLQELISIDEMELALNEIDNLTSEK---KIDQELAGVGNVDS-----574

QY 334 YLRPQWLEKEIEBEATKGLGPKFVIGVHVRRTDK--VGTEAAPHPIEYVMVHVEHFQ 391  
DB 575 -----GLRQLEERQKELEDEAPARGHEMQOLEANVASIRNELSHIPGYSQH-----K 622

QY 392 LLARMQVDKRVYATDPTLLKEAKTKYSNYEFISDNSISWSAGLHNRYTEN 445  
DB 623 MLRERLEAVEKRTAAKSLMSLRKTE----IDYEDIKTESIRLOEYNNMLLTN 672

RESULT 14  
S22695  
C:Species: Homo sapiens (man)  
C:Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 05-Nov-1999  
C:Accession: S22695; S19853  
R:Balbe, G.; Delabie, J.; Brueggen, J.; Richener, H.; Asselbergs, F.A.M.; Cerletti, N.;  
EMBO J. 11, 2103-2113, 1992  
A:Title: Restin: a novel intermediate filament-associated protein highly expressed in th  
A:Reference number: S22695; PMID:92289675; PMID:1600942  
A:Accession: S22695  
A:Molecule type: mRNA  
A:Residues: 1-1427 <BI>  
A:Cross-references: EMBL:X64838; NID:g35998; PIDN:CAA46050.1; PID:g35999  
C:Keywords: cytoskeleton

Query Match 3.7%; Score 115.5; DB 2; Length 1427;  
Best Local Similarity 17.9%; Pred. No. 17;  
Matches 101; Conservative 97; Mismatches 208; Indels 157; Gaps 20;

QY 40 SRELSKILAKLERLKQONEDL-----RRMAESLRIPGPDQGTATGRVR 84  
DB 588 SKENESLSKLESHANKENSVDIALWKSLETAISHQQAMEELKVSFSK-GLGTETAFA 646

QY 85 VLEEQVLKAK-----EQIENYKQARNGLKGKHHEILRRRIENGAKELWFFLOS 132  
DB 647 ELKTQIEKRLDYQHEIENLQNDQSDSRAHAKEAMEALRAKLMKVIKENSLEAIRSKL 706

QY 133 -----ELKLLKHLEG--NELQRHADEILLDLGHHSRIMTDL 167  
DB 707 DKAEDQHLVEMEDTNLQKQAEIKVKEILEVQAKCNEQTVIDNFTSOLKATEBKLL-DL 765

QY 168 YLISQTDGAGDREKEAKDLTELVRRTYLNQPK--DCSKARKL-----VCNIN 215

Db 766 DALRKASSEGKSEMKKLRQOLEAAAKQIKHLEIEKNAESSKASSITRELOGRELKLTNLQ 825

QY 216 KCGYGCQQLHHVVYCFMIAVGTQRTLLILESQWRVYATGWETVFRVSEFTCTDRSGLSTG 275

Db 826 ENLSEVSVQKTELL-----EKELQILKEKFAEASEEAVSVQSRMQET-----866

QY 276 HWSGEVNDKNIQVVELPIVDSLHPRPPYLP-----AVPEDLADRLLRVHGDPVAVWVS 329

Db 867 -----VNKLHQKESQFNMLSSDLEKLEENLADMEAKPREK-----D 902

QY 330 QVYLYLRPQWLEKEIEEATK-----KLGFKHPVIGVHVRRTDKVGTGTEAA 375

Db 903 EREBQILKAKELNDIAEIKWMSGDSSQITKWNDELRLKERDVEELQILTKANENAS 962

QY 376 F--HPIEEYVMVHVEHFQLLARRMQVDKRVYATDPTLLKEAKTKYSNYEFIS---DN 430

Db 963 FLOKSIEDMTVYKASQSQEAAKHEEKELEKRLSD--LEKKMETSHNQCEQLKARYER 1020

QY 431 SISWSAGLHNRY-----TENSUR-----GVILDIHFL-SQADPLVCTFSSQVC 472

Db 1021 ATSETTKTTHBIQLNQKTLTDTEDKLKGAREENSGLLQLEELRQAD---RAKAAQTA 1077

QY 473 RVAYEIMQTLHPDASANFHSILD 495

Db 1078 EDAMQIMEQMTKTKETTLASLED 1100

RESULT 15  
T19414  
hypothetical protein C23H4.6 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T19414  
R:Wilkinson, J.  
submitted to the EMBL Data Library, August 1996  
A:Reference number: Z19121  
A:Accession: T19414  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1137 <WIL>  
A:Cross-references: EMBL:Z78416; PIDN:CAB01681.1; GSPDB:GN00028; CESP:C23H4.6  
A:Experimental source: clone C23H4  
C:Genetics:  
A:Gene: CESP:C23H4.6  
A:Map position: X  
A:Introns: 34/1, 76/1, 111/2, 482/2, 598/2, 1091/3

Query Match 3.7%; Score 115; DB 2; Length 1137;  
Best Local Similarity 19.4%; Pred. No. 13;  
Matches 88; Conservative 75; Mismatches 156; Indels 134; Gaps 20;

QY 30 VRDNDHPDSSR-ELSKILAKL-----ERLKQONEDLRMAESLRIPGPDQGTAT 80  
DB 308 VKDAYKRTSRSEVQKKQLQELRDEVEVQDEETKESRADVNRKIVNDLKHEIKMSQ 367

QY 81 GRVRVLEBQLVKAKEIENYKQARNGLKG--DHEILRRRIENGAKELWFFLOSELKKLKH 139

Db 368 MRKVKVEIMLI--RKETAQAKQQLREALGKFGHELTTKLAE--AEDKRESLNIEIEIEH 424

QY 140 LEGNSLQRADEIILLDLGHHS-----RSIMTDLYYL-----SQTQDAGDWR 180

Db 425 VQLKALRKYKEKTKELRNEBEEKFNTRGKIATLRRTIEQDQKILRSMAKTKKNDVNFKG 484

QY 181 EKEAKDLTELVRRTYLNQPKDCSKARKLVNINKGCGYQOLHHVVYCFMIAVGTQRT 240

Db 485 PYMSILTEIHRSSQFKQPKG-----EBCFKMIANNFLCCSQEDATLR-----PLGKYVT 514

QY 241 TLILESQWRVYATGWETVFRVSE--CTDRSGLSTGHWSGEVNDKNIQVVELPIVDSLH 298

Db 515 LV--EPKWACAT--EBCFKMIANNFLCCSQEDATLR-----KIFDILKIPSD---559

QY 299 PRPPYLPVAPVEDLADRLLRVHGDPVAVWVVSQF-VKYLIRPQWLEKEIEEATKGLGFK 356



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 08:39:39 ; Search time 10.5 Seconds  
(without alignments)  
2575.272 Million cell updates/sec

Title: US-09-971-773-24

Perfect score: 3085

Sequence: 1 MRAWTGSRWIMLILFWGT.....YKREKIEIVKYPTPAEK 575

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt 41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	3074	99.6	575	1 FUT8_MOUSE	Q9wtc2 m alpha-(1,
2	3003	97.3	575	1 FUT8_HUMAN	Q9byc5 h alpha-(1,
3	2961	96.0	575	1 FUT8_BOVIN	Q9n0w2 b alpha-(1,
4	2910	94.3	575	1 FUT8_PIG	P79282 s alpha-(1,
5	126	4.1	997	1 SCPI_RAT	Q03410 rattus norv
6	125.5	4.1	845	1 SCPI_MESAU	Q60563 mesocricetu
7	120.5	3.9	751	1 COGS_DROME	Q9vj3d drosophila
8	119.5	3.9	1976	1 MYHA_BOVIN	Q27991 bos taurus
9	118.5	3.8	976	1 SCPI_HUMAN	Q15431 homo sapien
10	118	3.8	1727	1 ALM1_SCHPO	Q9utk5 schizosach
11	117	3.8	1955	1 PUMA_PARUN	Q61308 parascaris
12	116.5	3.8	919	1 RASO_AERPE	Q9yfi2 aeropyrum p
13	116.5	3.8	1969	1 MYGA_CAEEL	P12844 caenorhabdi
14	115.5	3.7	1427	1 REST_HUMAN	P30622 homo sapien
15	114	3.7	214	1 OSPI_HUMAN	Q92882 homo sapien
16	113	3.7	879	1 RASO_SULTO	Q95yr5 staphylococ
17	113	3.7	1057	1 CARB_STAAR	P58940 staphylococ
18	113	3.7	1217	1 ITN1_RAT	Q9wve9 rattus norv
19	113	3.7	3210	1 CENF_HUMAN	P49454 homo sapien
20	112.5	3.6	1270	1 ITN1_XENLA	Q42827 xenopus lae
21	112	3.6	1014	1 NEBL_HUMAN	Q76041 homo sapien
22	111.5	3.6	1278	1 DYNA_HUMAN	Q14203 homo sapien
23	111.5	3.6	1281	1 DYNA_MOUSE	Q08788 mus musculu
24	111.5	3.6	1935	1 MYH9_HUMAN	P12883 homo sapien
25	111.5	3.6	1961	1 MYH9_RAT	Q62812 rattus norv
26	111	3.6	993	1 SCPI_MOUSE	Q62209 mus musculu
27	111	3.6	1597	1 CRO_MOUSE	P49025 mus musculu
28	110.5	3.6	978	1 RASO_AQUAE	Q67124 aquifex aco
29	110	3.6	1057	1 CARB_STAAR	Q99ur5 staphylococ
30	110	3.6	1280	1 DYNA_RAT	P28023 rattus norv
31	109.5	3.5	1935	1 MYH7_RAT	P02564 rattus norv
32	109	3.5	352	1 SH32_MOUSE	Q62420 mus musculu
33	109	3.5	1230	1 UGS4_SOLTU	Q43846 solanum tub

#### RESULT 1

FUT8\_MOUSE  
ID FUT8\_MOUSE STANDARD; PRT; 575 AA.  
AC Q9WT52; Q921UL;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Alpha-(1,6)-fucosyltransferase (EC 2.4.1.68) (Glycoprotein 6-alpha-L-fucosyltransferase)  
DE fucosyltransferase (GDP-fucose--glycoprotein fucosyltransferase)  
DE (GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alpha1,6-fucosyltransferase)  
DE (alpha1-6FucT) (Fucosyltransferase 8).  
GN FUT8.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_taxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=20358720; PubMed=10902914;  
RA Hayaashi H., Yoneda A., Asada M., Ikeita M., Imamura T.;  
RT "Molecular cloning of mouse alpha-1,6-fucosyltransferase and expression of its mRNA in the developing cerebrum.";  
RL DNA Seq. 11:91-96 (2000).  
[2]

SEQUENCE FROM N.A.  
MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalios D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
CC -!- FUNCTION: Catalyzes the addition of fucose in alpha 1-6 linkage to the first GlcNAc residue, next to the peptide chains in N-glycans (By similarity).  
CC -!- CATALYTIC ACTIVITY: GDP-L-fucose + N(4)-[N-acetyl-beta-D-glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)]-[N-acetyl-beta-D-glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1->4)-N-acetyl-beta-D-glucosaminyl-(1->4)-N-acetyl-beta-D-glucosaminyl-(1->4)-N-acetyl-beta-D-glucosaminyl]asparagine = GDP + N(4)-[N-acetyl-beta-D-glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)]-[N-acetyl-beta-D-

glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1->6)]-  
>4)-N-acetyl-beta-D-glucosaminyl-(1->4)-[alpha-L-fucosyl-(1->6)]-  
N-acetyl-beta-D-glucosaminyl]asparagine.  
-!- PATHWAY: Glycosylation.  
-!- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound  
form in trans cisternae of Golgi (by similarity).  
-!- SIMILARITY: Contains 1 SH3 domain.  
-!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 23.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AB025198; BAA76392.1; -;  
DR EMBL; BC010666; AAH10666.1; -;  
DR MGI; MGI:1858901; Fut8.  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF00018; SH3; 1.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS0002; SH3; FALSE NEG.  
KW Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;  
KW Golgi stack; SH3 domain.  
FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 10 30 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
(POTENTIAL).  
FT DOMAIN 31 575 LUMENAL, CATALYTIC (POTENTIAL).  
FT DOMAIN 502 563 SH3.  
FT SITE 299 305 SH3-BINDING (POTENTIAL).  
FT SITE 365 366 IMPORTANT FOR DONOR-SUBSTRATE BINDING.  
FT CONFLICT 40 40 T -> S (IN REF. 2).  
FT CONFLICT 388 388 Q -> E (IN REF. 2).  
FT CONFLICT 418 418 N -> K (IN REF. 2).  
SQ SEQUENCE 575 AA; 66555 MW; 7BE2ED3146E0C45F CRC64;  
  
Query Match 99.6%; Score 3074; DB 1; Length 575;  
Best Local Similarity 99.5%; Pred. No. 6.9e-203;  
Matches 572; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 MRAWTGSRWIMLILPAWGTLFYIGCHLVDRNDHPDHSRSLKILAKLERLKQONEDL 60  
Db 1 MRAWTGSRWIMLILPAWGTLFYIGCHLVDRNDHPDHSRSLKILAKLERLKQONEDL 60  
QY 61 RMAESLRIPEDIGDTATGRVRLVEQLVKAKEIQENYKQARNGLGKDHILLRRRIE 120  
Db 61 RMAESLRIPEDIGDTATGRVRLVEQLVKAKEIQENYKQARNGLGKDHILLRRRIE 120  
QY 121 NGAKELWFFLQSELKLEHGNELQORHADEILLDGHHERSIMTDLVYLSOTDAGDWR 180  
Db 121 NGAKELWFFLQSELKLEHGNELQORHADEILLDGHHERSIMTDLVYLSOTDAGDWR 180  
QY 181 EKEAKDLTVORITLQNPDCSKARKLVNKNKGGYQGLHVVYCFMAYGTQRT 240  
Db 181 EKEAKDLTVORITLQNPDCSKARKLVNKNKGGYQGLHVVYCFMAYGTQRT 240  
QY 241 LILESQNWRYATGGWETVPRVSETCTDRSGLSGTGHSGEVNDKNIQVVELPIVDSLHP 300  
Db 241 LILESQNWRYATGGWETVPRVSETCTDRSGLSGTGHSGEVNDKNIQVVELPIVDSLHP 300  
QY 301 PPVPLPAVEDLADRLLRVHGDPVWVVSQFVKVLRPOPVEKEIEEATKLGKHPV 360  
Db 301 PPVPLPAVEDLADRLLRVHGDPVWVVSQFVKVLRPOPVEKEIEEATKLGKHPV 360  
QY 361 GVHVRRTDKVGTAEAFPHIEEYVHVVEEHFQLLARMQVDKRVYLATDPTLLKEATK 420  
Db 361 GVHVRRTDKVGTAEAFPHIEEYVHVVEEHFQLLARMQVDKRVYLATDPTLLKEATK 420  
QY 421 YSNYEFISNSISWSAGLHNRVTENSLRGVILDIHFLSQADFLVCTFSSQVCRAVEIMQ 480  
Db 421 YSNYEFISNSISWSAGLHNRVTENSLRGVILDIHFLSQADFLVCTFSSQVCRAVEIMQ 480

QY 481 TLHPDASANFHSLLDDIYFGQNAHNQIAVPHKPRTEEBIPMEPGDIIGVAGNHWGYS 540  
Db TLHPDASANFHSLLDDIYFGQNAHNQIAVPHKPRTEEBIPMEPGDIIGVAGNHWGYS 540  
QY 541 KGINKRLGKTGLPSYKVKREKTIETVKYPTYPEAEK 575  
Db 541 KGINKRLGKTGLPSYKVKREKTIETVKYPTYPEAEK 575

## RESULT 2

FUT8\_HUMAN  
ID FUT8\_HUMAN STANDARD; PRT; 575 AA.  
AC Q9BYC5; O00235; Q9BYC6; Q9P2U5; Q9P2U6;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Alpha-(1,6)-fucosyltransferase (EC 2.4.1.68) (Glycoprotein 6-alpha-L-  
fucosyltransferase) (GDP-fucose--glycoprotein fucosyltransferase)  
DE (GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alpha1,6-fucosyltransferase)  
DE (alpha1-6FucT) (Fucosyltransferase 8).  
GN FUT8.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1), AND FUNCTION.  
RX MEDLINE=97279058; PubMed=9133635;  
RA Yanagidani S., Uozumi N., Ihara Y., Miyoshi E., Yamaguchi N.,  
Taniguchi N.;  
RT "Purification and cDNA cloning of GDP-L-Fuc:N-acetyl-beta-D-  
glucosaminide:alpha1-6 fucosyltransferase (alpha1-6 FucT) from human  
gastric cancer MKN45 cells.";  
RL J. Biochem. 121:626-632(1997).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Embryo;  
RA Cailliau A., Balanzino L., Candellier J.J., Oriol R., Mollicone R.;  
RT "Differential splice variants of human FUT8 embryonic cDNA.";  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBAJ databases.  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RX MEDLINE=20275614; PubMed=10814706;  
RA Yamaguchi Y., Ikeda Y., Takahashi T., Ihara H., Tanaka T., Sasho C.,  
Uozumi N., Yanagidani S., Inoue S., Fujii J., Taniguchi N.;  
RT "Genomic structure and promoter analysis of the human  
alpha1,6-fucosyltransferase gene (FUT8).";  
RL Glycobiology 10:637-643(2000).  
RN [4]  
DONOR SUBSTRATE BINDING, AND MUTAGENESIS OF ARG-365 AND ARG-366.  
RX MEDLINE=20229550; PubMed=10764839;  
RA Takahashi T., Ikeda Y., Tateishi A., Yamaguchi Y., Ishikawa M.,  
Taniguchi N.;  
RT "A sequence motif involved in the donor substrate binding by  
alpha1,6-fucosyltransferase: the role of the conserved arginine  
residues.";  
RL Glycobiology 10:503-510(2000).  
CC -!- FUNCTION: Catalyzes the addition of fucose in alpha 1-6 linkage to  
the first GlcNAc residue, next to the peptide chains in N-glycans.  
CC -!- CATALYTIC ACTIVITY: GDP-L-fucose + N(4)-[N-acetyl-beta-D-  
glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)]-[N-acetyl-beta-D-  
glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1-  
>4)-N-acetyl-beta-D-glucosaminyl-(1->4)-[N-acetyl-beta-D-  
glucosaminyl]asparagine = GDP + N(4)-[N-acetyl-beta-D-  
glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)]-[N-acetyl-beta-D-  
glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1-  
>4)-N-acetyl-beta-D-glucosaminyl-(1->4)-[alpha-L-fucosyl-(1-  
>6)]-N-acetyl-beta-D-glucosaminyl]asparagine.  
CC -!- PATHWAY: Glycosylation.  
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound  
form in trans cisternae of Golgi (by similarity).  
CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=Q9BYC5-1; Sequence=Displayed;  
 CC Name=2; Synonyms=Retinal;  
 CC IsoId=Q9BYC5-2; Sequence=VSP 001807, VSP\_001808;  
 CC -!- SIMILARITY: Contains 1 SH3 domain.  
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 23.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; D89289; BAAL9764.1; -;  
 CC EMBL; AF038281; AAB92372.2; -;  
 CC EMBL; AF038280; AAB92372.2; JOINED.  
 CC EMBL; Y17979; CAA76988.1; -;  
 CC EMBL; Y17976; CAA76985.1; -;  
 CC EMBL; Y17977; CAA76986.1; -;  
 CC EMBL; Y17978; CAA76987.1; -;  
 CC EMBL; AB049828; BAB40975.1; -;  
 CC EMBL; AB049740; BAB40929.2; -;  
 CC EMBL; AB032573; BAA92859.2; -;  
 CC EMBL; AB032572; BAA92858.1; -;  
 CC EMBL; AB032568; BAA92858.1; JOINED.  
 CC EMBL; AB032569; BAA92858.1; JOINED.  
 CC EMBL; AB032570; BAA92858.1; JOINED.  
 CC EMBL; AB032571; BAA92858.1; JOINED.  
 CC PIR; JC5432; JC5432.  
 CC Genew; HGNC:4019; FUT8.  
 CC MIM; 602589; -;  
 CC InterPro; IPR001452; SH3.  
 CC Pfam; PF00018; SH3; 1.  
 CC SMART; SM00326; SH3; 1.  
 CC PROSITE; P55002; SH3; FALSE NEG.  
 CC Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;  
 KW Golgi stack; SH3 domain; Alternative splicing.  
 FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 10 30 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT LUMENAL, CATALYTIC (POTENTIAL).  
 FT SH3.  
 FT SH3-BINDING (POTENTIAL).  
 FT IMPORTANT FOR DONOR SUBSTRATE BINDING.  
 FT EVKDKNQVQVPELVIDSLHPRPPYLPVAPVEDLADRLVRVH  
 FT GDPVAVWVS -> TPIMNLLVITLFPGLDCTIDTQKIHV  
 FT E (in isoform 2).  
 FT /FTid=VSP 001807.  
 FT Missing (in isoform 2).  
 FT /FTid=VSP 001808.  
 FT R->A,K: COMPLETE LOSS OF ACTIVITY.  
 FT R->A,K: DECREASES ACTIVITY TO 3%.  
 FT SEQUENCE 575 AA; 66515 MW; 5A24A93881E18D0 CRC64;  
 SQ  
 Query Match 97.3%; Score 3003; DB 1; Length 575;  
 Best Local Similarity 96.5%; Pred. No. 5e-198;  
 Matches 555; Conservative 11; Mismatches 9; Indels 0; Gaps 0;  
 QY 1 MEAWTGSRWIMLILFANGTLLFYGGHLVRNDHPDHSRSLSKILAKLELKQONEDL 60  
 DB 1 MRPTWTGSRWIMLILFANGTLLFYGGHLVRNDHPDHSRSLSKILAKLELKQONEDL 60  
 QY 61 RMAESLRPEGPIDQGTATGRVRLVEQLVAKAEIQENYKKAQNGKGLKDEILRRRIE 120  
 DB 61 RMAESLRPEGPIDQGTATGRVRLVEQLVAKAEIQENYKKAQNGKGLKDEILRRRIE 120  
 QY 121 NGAKELWFFLQSELKLEKLEQLRHADEILDLGHHRSIMTDLVYLSQTDGAGDWR 180  
 DB 121 NGAKELWFFLQSELKLEKLEQLRHADEILDLGHHRSIMTDLVYLSQTDGAGDWR 180

QY 181 EKEAKDLTELQRRITVYLNQPKDCSKARKLVNINKGCGYQCLHVVVYCFMIAYGTORT 240  
 DB 181 EKEAKDLTELQRRITVYLNQPKDCSKARKLVNINKGCGYQCLHVVVYCFMIAYGTORT 240  
 QY 241 LILESQWRYATGWEVTFVRVSETCTDRSLGSHSGEYNDKNIQVVELPIVDSLHPR 300  
 DB 241 LILESQWRYATGWEVTFVRVSETCTDRSLGSHSGEYNDKNIQVVELPIVDSLHPR 300  
 QY 301 PPYLPVAPVEDLADRLRLVRHGDPAVWVSQFVKVLRPQWLEKEIEBATKLGPKHPVI 360  
 DB 301 PPYLPVAPVEDLADRLRLVRHGDPAVWVSQFVKVLRPQWLEKEIEBATKLGPKHPVI 360  
 QY 361 GVHVRTDKVTEAAPHPIEYVMVHESHFOLLARRMQVDKKRVYLATDDPTLLKEATK 420  
 DB 361 GVHVRTDKVTEAAPHPIEYVMVHESHFOLLARRMQVDKKRVYLATDDPTLLKEATK 420  
 QY 421 YSNTEFISDINSISWSAGLHNRYSRGLVLDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
 DB 421 YPNTEFISDINSISWSAGLHNRYSRGLVLDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
 QY 481 TLHPDASANFSLDDIYVFGGQNAHQIAVYPHPKPRTEEEIPMPBPDIIGVAGNHWGYS 540  
 DB 481 TLHPDASANFSLDDIYVFGGQNAHQIAVYAHQPRTADEIPMPBPDIIGVAGNHWGYS 540  
 QY 541 KGINKRLGKTGLYPSYKVKREKIETVKYPTYPEAEK 575  
 DB 541 KGVNRKLGRTGLYPSYKVKREKIETVKYPTYPEAEK 575  
 RESULT 3  
 FUT8\_BOVIN STANDARD; PRT; 575 AA.  
 AC Q9N0W2;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DE Alpha-(1,6)-fucosyltransferase (EC 2.4.1.68) (Glycoprotein 6-alpha-L-  
 DE fucosyltransferase) (GDP-fucose--glycoprotein fucosyltransferase)  
 DE (GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alpha1,6-fucosyltransferase)  
 DE (alpha1-6FucT) (Fucosyltransferase 8).  
 GN FUT8.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RX MEDLINE=20523969; PubMed=11070054;  
 RA Javard C., Dupuy F., Maftah A., Michalski J.-C., Oriol R.,  
 RA Petit J.-M., Julien R.;  
 RT "Ancestral exon organization of FUT8, the gene encoding the  
 RT alpha6-fucosyltransferase, reveals successive peptide domains which  
 RT suggest a particular three-dimensional core structure for the  
 RT alpha6-fucosyltransferase family";  
 RL Mol. Biol. Evol. 17:1661-1672(2000).  
 CC -!- FUNCTION: Catalyzes the addition of fucose in alpha 1-6 linkage to  
 CC the first GlcNAc residue, next to the peptide chains in N-glycans  
 CC (By similarity).  
 CC -!- CATALYTIC ACTIVITY: GDP-L-fucose + N(4)-{N-acetyl-beta-D-  
 CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)}-[N-acetyl-beta-D-  
 CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1-  
 CC >4)-N-acetyl-beta-D-glucosaminyl-(1->4)-N-acetyl-beta-D-  
 CC glucosaminyl]asparagine = GDP + N(4)-{N-acetyl-beta-D-  
 CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)}-[N-acetyl-beta-D-  
 CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1-  
 CC >4)-N-acetyl-beta-D-glucosaminyl-(1->4)-[alpha-L-fucosyl-(1->6)]-  
 CC N-acetyl-beta-D-glucosaminyl]asparagine.  
 CC -!- PATHWAY: Glycosylation.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound  
 CC form in trans cisternae of Golgi (By similarity).  
 CC -!- TISSUE SPECIFICITY: Highest expression found in brain. Also found

CC in heart, lung, spleen and kidney.  
 CC -!- SIMILARITY: Contains 1 SH3 domain.  
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 23.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----  
 CC EMBL: AF247186; AAF65460.1; -;  
 CC InterPro: IPR001452; SH3.  
 CC Pfam: PF00018; SH3; 1.  
 CC SMART: SM00326; SH3; 1.  
 CC PROSITE: PS50002; SH3; 1.  
 CC Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;  
 CC Golgi stack; SH3 domain.  
 CC DOMAIN 1 9  
 CC TRANSMEM 10 30  
 CC CYTOPLASMIC (POTENTIAL).  
 CC SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 CC (POTENTIAL).  
 CC LUMENAL, CATALYTIC (POTENTIAL).  
 CC SH3.  
 CC SH3-BINDING (POTENTIAL).  
 CC SH3-BINDING FOR DONOR SUBSTRATE BINDING (BY  
 CC SIMILARITY).  
 CC FT SEQUENCE 575 AA; 66492 MW; FE04C66B8A5BF540 CRC64;  
 CC  
 CC Query Match 96.0%; Score 2961; DB 1; Length 575;  
 CC Best Local Similarity 95.0%; Pred. No. 3.7e-195;  
 CC Matches 546; Conservative 15; Mismatches 14; Indels 0; Gaps 0;  
 CC  
 CC QY 1 MRAWTGSWIMILIFAWGTLFIYIGGHLVRDNDHSDHSSRELSKTLAKLERLKQONEDL 60  
 CC DB 1 MRPWTSWIMILIFAWGTLFIYIGGHLVRDNDHSDHSSRELSKTLAKLERLKQONEDL 60  
 CC QY 61 RMAESLRIPGIDOGTATGRVRLVEQLVKAKEQIENYKQARNGKLGKDHLELRRIE 120  
 CC DB 61 RMAESLRIPGIDOGTATGRVRLVEQLVKAKEQIENYKQARNGKLGKDHLELRRIE 120  
 CC QY 121 NGAKELWFFLOSELKXKLEGNELQRADEILLDLGHHSRIMTDLVYLSOTDGGDWR 180  
 CC DB 121 NGAKELWFFLOSELKXKLEGNELQRADEILLDLGHHSRIMTDLVYLSOTDGGDWR 180  
 CC QY 181 EKAEDLTQLVQRITVYLNQPKDCSKARKLVNCKNGCGYQGLHVVVCFMIAVGTQT 240  
 CC DB 181 EKAEDLTQLVQRITVYLNQPKDCSKARKLVNCKNGCGYQGLHVVVCFMIAVGTQT 240  
 CC QY 241 LILESQNRVYATGGWETVFRPVSETCTDRSGLSTGHSGEVNDKNIQVVELPIVDSLHPR 300  
 CC DB 241 LILESQNRVYATGGWETVFRPVSETCTDRSGLSTGHSGEVNDKNIQVVELPIVDSLHPR 300  
 CC QY 301 PVLPLAVPEDLADRLRLVHGDPAVWVVSQFVKYLIRPQWLEKEIEEATKLGKHPVI 360  
 CC DB 301 PVLPLAVPEDLADRLRLVHGDPAVWVVSQFVKYLIRPQWLEKEIEEATKLGKHPVI 360  
 CC QY 361 GVHVRITDKVGTAAFPPIIEYVHVVEHFQILLARMQVDDKRVYLATDPTLLKEAKTK 420  
 CC DB 361 GVHVRITDKVGTAAFPPIIEYVHVVEHFQILLARMQVDDKRVYLATDPTLLKEAKTK 420  
 CC QY 421 YSNYEFISDINSISWSAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
 CC DB 421 YSNYEFISDINSISWSAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
 CC QY 481 TLHPDASANPHSLDDIYIFGQNAHQIAVYHPKPTBEEIPMEPGDIIICVAGNWDGYS 540  
 CC DB 481 TLHPDASANPHSLDDIYIFGQNAHQIAVYHPKPTBEEIPMEPGDIIICVAGNWDGYS 540  
 CC QY 541 KGNRNLKGTGLYPSVKVREKIEYVYVYVPEAK 575  
 CC DB 541 KGNRNLKGTGLYPSVKVREKIEYVYVYVPEAK 575

RESULT 4  
 FUT8\_PIG ID FUT8\_PIG STANDARD; PRT; 575 AA.  
 AC P79282;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Alpha-(1,6)-fucosyltransferase (EC 2.4.1.68) (Glycoprotein 6-alpha-L-  
 DE fucosyltransferase) (GDP-fucose--glycoprotein fucosyltransferase)  
 DE (GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alpha1,6-fucosyltransferase)  
 DE (alpha1-6FucT) (Fucosyltransferase 8).  
 GN FUT8.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euthera; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A., SEQUENCE OF 102-130; 333-344; 402-415 AND  
 RP 566-575, AND FUNCTION.  
 RC TISSUE=Brain;  
 RX MEDLINE=97066976; PubMed=8910378;  
 RA Uozumi N., Yanagidani S., Miyoshi E., Ihara Y., Sakuma T., Gao C.-X.,  
 RA Teshima T., Fujii S., Shiba T., Taniguchi N.;  
 RT "Purification and cDNA cloning of porcine brain GDP-L-Fuc:N-acetyl-  
 RT beta-D-glucosaminide alpha1--6fucosyltransferase.";  
 RL J. Biol. Chem. 271:27810-27817(1996).  
 CC -!- FUNCTION: Catalyzes the addition of fucose in alpha 1-6 linkage to  
 CC the first GlcNAc residue, next to the peptide chains in N-glycans.  
 CC The pH optimum is 7.  
 CC -!- CATALYTIC ACTIVITY: GDP-L-fucose + N(4)-(N-acetyl-beta-D-  
 CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)-[N-acetyl-beta-D-  
 CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1-  
 CC >4)-N-acetyl-beta-D-glucosaminyl-(1->4)-N-acetyl-beta-D-  
 CC glucosaminyl]asparagine = GDP + N(4)-(N-acetyl-beta-D-  
 CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)-[N-acetyl-beta-D-  
 CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1-  
 CC >4)-N-acetyl-beta-D-glucosaminyl-(1->4)-[alpha-L-fucosyl-(1->6)]-  
 CC N-acetyl-beta-D-glucosaminyl]asparagine.  
 CC -!- PATHWAY: Glycosylation.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound  
 CC form in trans cisternae of Golgi (By similarity).  
 CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN BRAIN.  
 CC -!- SIMILARITY: Contains 1 SH3 domain.  
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 23.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----  
 CC EMBL: D86723; BAAL3157.1; -;  
 CC InterPro: IPR001452; SH3.  
 CC Pfam: PF00018; SH3; 1.  
 CC SMART: SM00326; SH3; 1.  
 CC PROSITE: PS50002; SH3; 1.  
 CC Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;  
 CC Golgi stack; SH3 domain.  
 CC DOMAIN 1 9  
 CC TRANSMEM 10 30  
 CC CYTOPLASMIC (POTENTIAL).  
 CC SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 CC (POTENTIAL).  
 CC LUMENAL, CATALYTIC (POTENTIAL).  
 CC SH3.  
 CC SH3-BINDING (POTENTIAL).  
 CC SH3-BINDING FOR DONOR SUBSTRATE BINDING (BY  
 CC SIMILARITY).  
 CC FT SEQUENCE 575 AA; 66229 MW; OF199D0BC2018F7B CRC64;  
 CC  
 CC Query Match 94.3%; Score 2910; DB 1; Length 575;  
 CC Best Local Similarity 93.6%; Pred. No. 1.1e-191;



Matches 538; Conservative 14; Mismatches 23; Indels 0; Gaps 0;

QY 1 MRATGSRWMLILFANGTLLFYGGHLVRDNDHDPDSSRELKILAKLERLKOQNDL 60  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 1 MRPTGSRWMLILFANGTLLFYGGHLVRDNDHDPDSSRELKILAKLERLKOQNDL 60  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 61 RMMAESLRIPGPIDQGTATGVRVLEBQLVKAKEQIENYKQARNGLKGDKHEILRRRIE 120  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 61 RMMAESLRIPGPIDQGTATGVRVLEBQLVKAKEQIENYKQARNGLKGDKHEILRRRIE 120  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 121 NGAKELWFFLOSELKGLKLEGNELORHADEILDLGHHERSIMTDLVYLSQDAGDWR 180  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 121 NGAKELWFFLOSELKGLKLEGNELORHADEILDLGHHERSIMTDLVYLSQDAGDWR 180  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 181 EKEAKDLTELVRRTIYLNQPKDCSKARKLVNKNKGGYGCQLHHVYVCFMIAVGTORT 240  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 181 EKEAKDLTELVRRTIYLNQPKDCSKARKLVNKNKGGYGCQLHHVYVCFMIAVGTORT 240  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 241 LILESQNRVYATGWTETFRPVSETCTDRSGLSLTHWSGEVNDKNIQVVELPIVDSLHPR 300  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 241 LALESHNRVYATGWTETFRPVSETCTDRSGLSLTHWSGEVNDKNIQVVELPIVDSLHPR 300  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 301 PYPPLAVPEDLADRLLRVHGDPVWVVSQFVKYLIRPQPMLEKEIEBATKGLGPKHPVI 360  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 301 PYPPLAVPEDLADRLLRVHGDPVWVVSQFVKYLIRPQPMLEKEIEBATKGLGPKHPVI 360  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 361 GVHVRTDKVTEAAAFHPIEEMVVEHFQILARRMQVDKRVYLATDPTLLKEATK 420  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 361 GVHVRTDKVTEAAAFHPIEEMVVEHFQILARRMQVDKRVYLATDPTLLKEATK 420  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 421 YSNRYFISDNTSWSAGLHNRVYTNENSLRGVILDIHFLSOADFLVCTFSQVCRVAYEIMQ 480  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 421 YPSYFISDNTSWSAGLHNRVYTNENSLRGVILDIHFLSOADFLVCTFSQVCRVAYEIMQ 480  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 481 TLHPDASANFSLDDIYVFGGQNAHQIAYVPHKPRTEEBEIPMEPGDIIGVAGNWDGYS 540  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 481 ALHPDASANFSLDDIYVFGGQNAHQIAYVPHKPRTEEBEIPMEPGDIIGVAGNWDGYP 540  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 541 KGINKLGTGLYPSKYREKLETYKPYTPAEK 575  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 541 KGVRKLGRTGLYPSKYREKLETYKPYTPAEK 575  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 5

SCPI\_RAT  
 ID SCPI\_RAT STANDARD; PRT; 997 AA.  
 AC Q03410;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Synaptonemal complex protein 1 (SCP-1 protein).  
 GN SCPI OR SCPI.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=9309984; PubMed=1464329;  
 RA Meuwissen R.L.J., Offenbergh H.H., Dietrich A.J., Riesewijk A.,  
 RA Iersel M., Heyting C.;  
 RT "A coiled-coil related protein specific for synapsed regions of  
 RT meiotic prophase chromosomes.";  
 RL EMBO J. 11:5091-5100(1992).  
 CC -!- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF  
 CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS  
 CC CHROMOSOMES DURING MEIOTIC PROPHASE.  
 CC SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF  
 CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.  
 CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS  
 CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE  
 CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE

SYNAPTONEMAL COMPLEX.  
 -!- TISSUE SPECIFICITY: Testis.  
 -!- DEVELOPMENTAL STAGE: EXPRESSED EXCLUSIVELY IN MEIOTIC PROPHASE  
 CELLS.  
 -!- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,  
 FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL  
 DOMAIN HAS DNA-BINDING CAPACITY.  
 -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS  
 CORRECTED IN POSITION 6 TO MAXIMIZE THE SIMILARITY WITH THE  
 OTHER SPECIES SCPI SEQUENCES.

-----  
 THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

EMBL; X67805; CAA4806.1; ALT FRAME.  
 Nuclear protein; Meiosis; Cell division; Phosphorylation;  
 DNA-binding; Coiled coil.  
 DOMAIN 108 819 COILED COIL (POTENTIAL).  
 DOMAIN 118 121 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 DOMAIN 701 704 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 DOMAIN 902 905 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 DOMAIN 982 990 ARG/LYS-RICH (BASIC).  
 SEQUENCE 997 AA; 116511 MW; 229059823FD684BE CRC64;

Query Match 4.1%; Score 126; DB 1; Length 997;  
 Best Local Similarity 20.6%; Pred. No. 0.65;  
 Matches 109; Conservative 78; Mismatches 180; Indels 162; Gaps 24;

QY 42 ELSKILAKLERLKOQNDLRRMAESLRIPGPIDQGTATGVRVLEBQLVKAKEQIENY 100  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 155 ENEKSVSLKLEIEIQENKDL-----IKENNATRWCMNLLKETCARSAEKTSKY 201  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 101 KKQ-----ARNGLGK--DHEILRRRIENGAKELWFFLOSELKGLKLEGNELQR 147  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 202 EYEREETQVYVLDNNIEKMLAFEEELRVAENARLEMHFKLKEDEKIQHLE-EEYQK 260  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 148 HAD-----EILDLGHHERSIMTDLVYL--SQTGAGDWRKEBAKDLTELVRRTIYQ 199  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 261 EVNNKENQVSLLIQSTKENKMKDLATFLLEESRDKANQLEEK-----TKLDENLKELN 315  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 200 NPKDCSRARKLVNKNKGGYGCQLHHVYVCFMIAVGTORTLILLESQNRVYATGWTETVF 259  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 316 EKDHLS-----ELEDIKMSMQRSMSQTKTL---EEDLQIAT---KTIY 354  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 260 RPVSE-----TCTDRSGLSLTHWSGEVND----- 283  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 355 QLTEEKAEQMEELNKAHTTHSLVTELTAKTCTLEELRTEQORLENEDQLKITMELQ 414  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 284 -----KNIQVVELPIVDSLHPRPYPPLAVPEDLADRLLRVHG-DPAVWVVSQ 330  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 415 KKSSELEBMTFKNKEVELEELKTI-----LAEDQKLDELKQVEKLAELQKBEQ 466  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 331 FVKYLIRPQPMLEKEIEBATKGLGPKHPVIGVHVRTDKVTEAAAFHPIEEMVVEH 388  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 467 ELTFELQTR---EKEIHDL-----VQVTYTKTSE-----EHLKQVEEMK 504  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 389 -----HFOLLARR---MOVDKRVYLATDPTLLKEAKTKYSNYEFTSDNISWSA 436  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 505 TELSEKELKNIELTANSOMLLENKLVQELASDMVLELKQKQEDINCKQOERMLKQIE 564  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 437 GLHNRVYATGWTETFRPVSETCTDRSGLSLTHWSGEVNDKNIQVVELPIVDSLHPR 480  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 565 TLEEK--EMNLRDELESVRKEFIQGGDEVKCKLDKSEENARSIEVLK 611  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 6  
 SCPI\_MESAU  
 ID SCPI\_MESAU STANDARD; PRT; 845 AA.

AC Q60563;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Synaptonemal complex protein 1 (SCP-1 protein) (Meiotic chromosome  
 DE synaptic protein) (Fragment).  
 GN SCPI OR SYNI.  
 OS Mesocricetus auratus (Golden hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Mesocricetus.  
 OX NCBI\_TaxID=10036;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=95181577; PubMed=7876343;  
 RA Dobson M.J., Pearlman R.E., Karaiskakis A., Spyropoulos B.,  
 RA Moens P.B.;  
 RT "Synaptonemal complex proteins: occurrence, epitope mapping and  
 RT chromosome disjunction";  
 RL J. Cell Sci. 107:2749-2760 (1994).  
 CC -!- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF  
 CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS  
 CC CHROMOSOMES DURING MEIOTIC PROPHASE (BY SIMILARITY). HAS NON-  
 CC SPECIFIC DNA BINDING CAPABILITY.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF  
 CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.  
 CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS  
 CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE  
 CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE  
 CC SYNAPTONEMAL COMPLEX.  
 CC -!- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,  
 CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL  
 CC DOMAIN HAS DNA-BINDING CAPACITY (BY SIMILARITY).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; L32978; AAC42039.1; --  
 DR PTR; 148176; 148176.  
 KW Nuclear protein; Meiosis; Cell division; Phosphorylation;  
 KW DNA-binding; Coiled coil.  
 FT NON\_TER 1  
 FT DOMAIN <1 672 COILED COIL (POTENTIAL).  
 FT DOMAIN 553 556 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 753 756 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 830 838 ARG/LYS-RICH (BASIC).  
 SQ SEQUENCE 845 AA; 99401 MW; D7F28873C824C6A8 CRC64;  
 Query Match 4.1%; Score 125.5; DB 1; Length 845;  
 Best Local Similarity 19.7%; Pred. No. 0.57;  
 Matches 85; Conservative 81; Mismatches 161; Indels 105; Gaps 17;  
 QY 42 ELKILAKLERLKOQNEEDLRMAESLRPEGPID--QGTATGVRVRLERQVAKAEQIEN 99  
 DB 288 ELKKILAEQKLDEKKQVEKLAELQGEQELTLLOTREREVHDLESQILVTKISDON 347  
 QY 100 YKQARNGLKGDHEILRRRIENGAKELWFFLOSELKKLKH-----LEGNELORH 148  
 DB 348 YSKQV-----ELKTKLEE-----ELKNAELTASCGLSLNNKLTQE 386  
 QY 149 ADEILDGLGHERSIWT----DIYLSQTDGADWREKAQDLTELIVQRRTIYLNPKDC 204  
 DB 387 TNDMALELKYQEDITNSKKQEBRLKQIENLEE-KETHLRDELSVRKEFIQOGNEVRC 445  
 QY 205 -----SKARKLVNINKGCGYCOLHHVYCFMAYGQRTILLESQ--NRYATGGWE 256  
 DB 446 KLDKSEBENARSIECVLK-----KEQMKILENCKNNLRKQA---E 483

## RESULT 7

COG5\_DROME  
 ID COG5\_DROME STANDARD; PRT; 751 AA.  
 AC Q9VJD3; Q8T0G3;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Putative conserved oligomeric Golgi complex subunit 5 (Four way stop  
 DE protein).  
 GN FWS OR CG6549.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bekeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang X., Lin X.,  
 RA Mearns K., McInerney J., McInerney T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kimms I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Voh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster."; Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE OF 406-751 FROM N.A.  
RX STRAIN=Berkeley; TISSUE=Embryo;  
MEDLINE=22426066; PubMed=12537569;  
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,  
Rubin G.M., Celnik S.E.;  
RT "A Drosophila full-length cDNA resource";  
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
CC -!- FUNCTION: Required for normal Golgi function (By similarity).  
CC -!- SUBUNIT: Component of the conserved oligomeric Golgi complex which  
CC is composed of eight different subunits and is required for normal  
CC Golgi morphology and localization (By similarity).  
CC -!- SUBCELLULAR LOCATION: Golgi (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE COG FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF003655; RA53619.1; -  
DR EMBL; AY069335; AL39480.1; ALT\_INIT.  
DR Flybase; FBgn0024689; fws.  
KW Hypothetical protein; Transport; Protein transport; Golgi stack;  
KW Membrane.  
SQ SEQUENCE 751 AA; 84953 MW; E32B1D05BEEF6EBA CRC64;  
  
Query Match 3.9%; Score 120.5; DB 1; Length 751;  
Best Local Similarity 20.7%; Pred. No. 1.1; Indels 177; Gaps 32;  
Matches 127; Conservative 97; Mismatches 213;  
  
QY 32 DNDPDHSS-----RELKILAKLERLKQONDLRMASLRPEGP-IDQGTATCR 82  
DB 19 DNDFTASMSHLITGOIQELSK-----QLQTKELH---QOVRDKGALLQOATHGR 69  
QY 83 ----VRVLEQLVKAKE-----QIENYKQARNG---LGKDHEILRRRIENGAKELWF 128  
DB 70 FDAALNALAEDQVRVRETHRLKNQVDQYQVENQTVLGRHDVSHLLRSAGT----- 124  
QY 129 FLOSELKLLKLEGNELQORHADEILDLGH-HRSIMTDLVYLSQTDGAGDREKADL 187  
DB 125 -LLSLTAKLKATK--DVLRLA-EIHPGLQIEDKELKDIDFIOQERAYVSSAQKIRNL 180  
QY 188 TELVQRITVLO--NEPKDCSKARKVCNIN-----KGCYGGCOL 224  
DB 181 TQM--QLVTGLQERNENQVNAKIFMNFNTLEKSLDNLATFIADMEQSLKECFAGNDI 238  
QY 225 H-----HVVCFMIAYGTORT-LILESQNRVYATGGWET-----VFRPVSETCTDRSL 272  
DB 239 SVLNKSPTHNVSKPAPSRGPKTPQTITQNR--AKFWKSLHLLYDELFTCTQIKLL 296  
QY 273 STG-----HNSGEVNDKNI-----QVVELPIVSLHPRPPYPLVAPVEDLADRL-- 317  
DB 297 KTALEQINGFGVTSSESDQICPQRFQVQQLLRKSFDECPQHVTQTLOEGLSKLLTSAR 356  
QY 318 ----RVHGDPAVWVSQFVKYLRPQWLEKEIBEATK-KLGFKHPVIGVHVRRYDKV-- 370  
DB 357 GLEQRLHGE-----PQFNELFAP-----LEVGVVSKAANFKACLAGVDLPFGNETVDN 405  
QY 371 ----GTEAAPHPIEYMHV-----BEHFOLLAR-RMQVDKKEVYLATDDPTLL 414  
DB 406 FIRVASTELSAALIDSLRLNANVFAACGKELCTKLEAQIKLGADSKQV---VDLPNLQ 462  
QY 415 KEAKTKYNYEFTSDNSISWSAGLHNRYTENSRLGVLDIHF-----LS 458  
DB 463 QOQNTQLANLVF-----YYKDSVRRMLSDLVHVFHEFTPGTAREIISRLSLE 507

QY 459 QADFLVCTFSQVCRVAYE-----IMQLHPDASANFHSLD-----DIYFGQNAH 505  
DB 508 QADLLIGTIIQQIMESIITTSIIIVLSMHREPGLNSERMSTTGPSPMYMKELQEFVNRWS 567  
QY 506 NQIAVYPHKPRTTE 519  
DB 568 HHIALFDKDK(MTKK 581  
  
RESULT 8  
MYHA\_BOVIN  
ID MYHA\_BOVIN STANDARD; PRT; 1976 AA.  
AC Q27991;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Myosin heavy chain, nonmuscle type B (Cellular myosin heavy chain,  
DE type B) (Nonmuscle, myosin heavy chain-B) (NMHC-B).  
GN MYH10.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ohara M., Ishiguro N., Shinagawa M.;  
RT "Bos taurus nonmuscle myosin heavy chain B mRNA, complete cds.";  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 204-302 FROM N.A.  
RC TISSUE=Brain cortex;  
RX MEDLINE=95301542; PubMed=7782316;  
RA Itoh K., Adelstein R.S.;  
RT "Neuronal cell expression of inserted isoforms of vertebrate nonmuscle  
RT myosin heavy chain II-B";  
RL J. Biol. Chem. 270:14533-14540(1995).  
CC -!- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,  
CC CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND  
CC CAPING (BY SIMILARITY).  
CC -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY  
CC CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2  
CC REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,  
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.  
CC -!- SIMILARITY: Contains 1 IQ domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AB022023; BAA16494.1; -  
DR EMBL; U15716; AAA87715.1; -  
DR HSSP; P10587; 1BR2.  
DR InterPro; IPR000048; IQ region.  
DR InterPro; IPR001609; myosin head.  
DR InterPro; IPR004009; Myosin\_N.  
DR InterPro; IPR002928; Myosin\_tail.  
DR InterPro; IPR002017; Spectrin.  
DR Pfam; PF00612; IQ; 1.  
DR Pfam; PF00063; myosin head; 1.  
DR Pfam; PF02736; Myosin\_N; 1.  
DR Pfam; PF01576; Myosin\_tail; 1.  
DR PRINTS; PR00193; MYOSINHEAVY.  
DR ProDom; PD000355; myosin\_head; 1.  
DR SMART; SM00015; IQ; 1.  
DR SMART; SM00242; MYSC; 1.

DR PROSITE; PS50096; IQ; 1.  
 KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;  
 FT Coiled coil; Alkylation; Multigene family.  
 FT DOMAIN 1 785 MYOSIN HEAD-LIKE.  
 FT DOMAIN 786 815 IQ.  
 FT DOMAIN 845 1976 COILED COIL (POTENTIAL).  
 FT NP BIND 178 185 ATP (POTENTIAL).  
 FT MOD RES 701 711 ALKYLATION (SH-1) (POTENTIAL).  
 FT MOD RES 711 711 ALKYLATION (SH-2) (POTENTIAL).  
 SQ SEQUENCE 1976 AA; 22909 MW; 6144354451C0F790 CRC64;

Query Match 3.9%; Score 119.5; DB 1; Length 1976;  
 Best Local Similarity 25.7%; Pred. No. 4.5;  
 Matches 45; Conservative 33; Mismatches 44; Indels 53; Gaps 8;

QY 41 RELSKILAKLERLKKQNEEDLRMAESLRIPGPDQGTATGRVRLVEQLVKAQEQ---96  
 DQ 1339 RQLEERSSLOEQEESSEARRLEK-----QLQALQALQDTKKKVVDD 1383  
 QY 97 ---IENYKKQARNGLKDHILRRRIENGAKELWFFLQSELKHLKLEGNELQHADEL 153  
 DQ 1384 LGTIENL-EAEEKKLLKDEVLQRLEKA-----LAYDKLETK---TRLQQLDLDL 1433  
 QY 154 LDIGHERSIMTDL-----YYLSQTDGA-GDWREKAKDLT 188  
 DQ 1434 VDL-DHQRTQVSNLEKKQKKFDQLLABEKNISARYAEERDRAEAEREKETRLS 1487

## RESULT 9

SCPI\_HUMAN STANDARD; PRT; 976 AA.  
 ID SCPI\_HUMAN  
 AC Q15431; O14963;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Synaptonemal complex protein 1 (SCP-1 protein).  
 GN SYCP1 OR SCPI.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=97224467; PubMed=911375;  
 RA Meuwissen R.L.J., Meerts I., Hoovers J.M.N., Leschot N.J.,  
 RA Heyting C.;  
 RT "Human synaptonemal complex protein 1 (SCPI): isolation and  
 RT characterization of the cDNA and chromosomal localization of the  
 RT gene.";  
 RL Genomics 39:377-384(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=98037449; PubMed=9371398;  
 RA Kondoh N., Nishina Y., Tachida J., Koga M., Tanaka H., Uchida K.,  
 RA Inazawa J., Taketo M., Nozaki M., Nojima H., Matsumiya K., Namiki M.,  
 RA Okuyama A., Nishimura Y.;  
 RT "Assignment of synaptonemal complex protein 1 (SCPI) to human  
 RT chromosome ip13 by fluorescence in situ hybridization and its  
 RT expression in the testis.";  
 RL Cytogenet. Cell Genet. 78:103-104(1997).  
 CC -1- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF  
 CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS  
 CC CHROMOSOMES DURING MEIOTIC PROPHASE.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF  
 CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.  
 CC IS FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS  
 CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE  
 CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE  
 CC SYNAPTONEMAL COMPLEX (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: Testis.  
 CC -1- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,

CC CC  
 CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL  
 CC DOMAIN HAS DNA-BINDING CAPACITY (BY SIMILARITY).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X95654; CAA64956.1; -;  
 DR EMBL; D67035; BAA22586.1; -;  
 DR Genbank; HGNC:11487; SYCP1.  
 DR MIM; 602162; -;  
 DR GO; GO:0005634; C:nucleus; TAS.  
 DR GO; GO:0005716; C:synaptonemal complex; TAS.  
 DR GO; GO:0003677; F:DNA binding activity; TAS.  
 DR GO; GO:0007131; F:meiotic recombination; TAS.  
 DR GO; GO:0007283; P:spermatogenesis; TAS.  
 DR GO; GO:0007129; P:synapsis; TAS.  
 KW Nuclear protein; Meiosis; Cell division; Phosphorylation;  
 KW DNA-binding; Coiled coil.  
 FT DOMAIN 12 100 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 107 798 COILED COIL (POTENTIAL).  
 FT DOMAIN 117 120 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 679 682 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 880 883 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 961 969 ARG/LYS-RICH (BASIC).  
 FT CONFLICT 46 46 L -> F (IN REF. 2).  
 FT CONFLICT 106 106 F -> Y (IN REF. 2).  
 FT CONFLICT 153 153 F -> C (IN REF. 2).  
 FT CONFLICT 161 161 K -> T (IN REF. 2).  
 FT CONFLICT 168 168 E -> D (IN REF. 2).  
 FT CONFLICT 216 216 N -> S (IN REF. 2).  
 FT CONFLICT 225 226 HG -> FE (IN REF. 2).  
 FT CONFLICT 350 350 K -> N (IN REF. 2).  
 FT CONFLICT 360 360 E -> D (IN REF. 2).  
 FT CONFLICT 400 401 KN -> NY (IN REF. 2).  
 FT CONFLICT 406 406 K -> I (IN REF. 2).  
 FT CONFLICT 415 415 K -> T (IN REF. 2).  
 FT CONFLICT 449 449 E -> D (IN REF. 2).  
 FT CONFLICT 483 510 IQLTAITTSQYKSEKVDKTELENEK -> YSYCHYHKV  
 FT CONFLICT 516 528 TVLPKRGQRPKLSKRE (IN REF. 2).  
 FT CONFLICT 549 549 N -> I (IN REF. 2).  
 FT CONFLICT 560 560 K -> T (IN REF. 2).  
 FT CONFLICT 805 805 E -> D (IN REF. 2).  
 FT CONFLICT 941 941 P -> S (IN REF. 2).  
 SQ SEQUENCE 976 AA; 114069 MW; 8BA81D042AC2696B CRC64;

Query Match 3.8%; Score 118.5; DB 1; Length 976;  
 Best Local Similarity 19.5%; Pred. No. 2.1; Indels 139; Gaps 18;  
 Matches 91; Conservative 66; Mismatches 171;

QY 45 KILAKLERLKKQNEEDLRMAESLRIPGPDQGTATGRV-RVLEEQLVKAKEIQENYKKQ 103  
 DQ 157 KVSLEKLEGIQENKDL-----IKENNAIRHLNLLKTCARSAEKTKKYEVE 203  
 QY 104 -----ARNGLKG---DHEILRRRIENGAKELWFFLQSELKHLKLEGNELQHADEL 144  
 DQ 204 REETQVYMDLNNNIEKMITAHGELRVQAENSRLEMHFKLAKEDYEKIQLHLEQYKKEIND 263  
 QY 145 LQRADEITLLDGHHSIMTDLVYL---SQTDGADWRE-----K 182  
 DQ 264 KEQVSLDILQITSEKKN-MKDLTFLEESRDKVNQLEKTKLOSENKQSKIEKQHLLTK 322  
 QY 183 EAKOLTELIVQRRRI-TYLNQPKDCSKARKLVNINKGCGYGCQLHHVVYCFMAYGTQRTL 241  
 DQ 323 ELEDIKVSLQRSVSTQKALEEDLQIATKTIQOLTE-----EKETQ 362  
 QY 242 ILESQNMRYATGGWETVFRPVSETCTDRSGLSLSTGHWSGEVNDKNIQVVELFVDSLHPRP 301

```

Db 363 MEESKARAAGSFVVTEPE -TTVCSELLRTEQORLEKNEKDQLKILTMELQKK----- 415
Qy 302 PVLPLAVPEDLADRLRLVHGDPVWVVSQFVYLIRPQPWLEKEIEEATKLGFKHPVIG 361
Db 416 -----SSELEEMTKLTNNKEVELEELKVLGKERTLLY 448
Qy 362 VHVRRTDKVGTEAAFPHEEYVMVVEHP -OLLARRMOVDKRVVL---ATDDPTLLKE 416
Db 449 ENKQ-----PEKTAELKGTQEOLIGLLQAREKVEHDELTQLTAITTSQYYSKE 498
Qy 417 AK-----TKYSNYEFISD-NSISWSAGLHNRVTENSLRGVILDI 454
Db 499 VKDLKTELENEKLNKLTSHCNKLS-----LENKELTQETSDMTLEL 541

RESULT 10
ALM1 SCHPO
ID ALM1 SCHPO STANDARD; PRT: 1727 AA.
AC Q9UTK5; O13313; O9UT78;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Abnormal long morphology protein 1 (Sp8).
GN ALM1 OR SPAC1486.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN (1) _SEQUENCE FROM N.A.
RP STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitz E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Foreburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrall B.G., Nurse P.;
RA "The genome sequence of Schizosaccharomyces pombe";
RL Nature 415:871-880(2002).
[2]
RP SEQUENCE OF 495-1727 FROM N.A., AND CHARACTERIZATION.
RC STRAIN=972;
RX MEDLINE=20123449; PubMed=10660053;
RA Jimenez M., Petit T., Gancedo C., Goday C.;
RT "The alm1+ gene from Schizosaccharomyces pombe encodes a coiled-coil
RT protein that associates with the medial region during mitosis.";
RL Mol. Gen. Genet. 262:921-930(2000).
[3]
RN
RP SEQUENCE OF 644-834 FROM N.A.
RC STRAIN=968 h90;
RX MEDLINE=20223868; PubMed=10759889;
RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,

```

```

RA Hiraoka Y.;
RT "Large-scale screening of intracellular protein localization in living
RT fission yeast cells by the use of a GFP-fusion genomic DNA library.";
RL Genes Cells 5:169-190(2000).
CC -!- FUNCTION: AFTER THE ONSET OF MITOSIS, AT MID- TO LATE ANAPHASE,
CC CO-LOCALIZES WITH THE MEDIAL ACTIN RING. MAY PLAY A ROLE IN
CC CYTOKINESIS.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL133357; CAJ62414.1; -
CC EMBL; AF010473; BAA65416.1; ALT_INIT.
CC EMBL; AB028012; BAA87316.1; -
CC PIR; T50073; T50073.
CC GeneDB SPombe; SPAC1486.04c; -
CC Coiled coil. 57 361 COILED COIL (POTENTIAL).
CC DOMAIN 443 463 COILED COIL (POTENTIAL).
CC DOMAIN 542 740 COILED COIL (POTENTIAL).
CC DOMAIN 804 1106 COILED COIL (POTENTIAL).
CC DOMAIN 1223 1427 COILED COIL (POTENTIAL).
CC DOMAIN 1497 1555 COILED COIL (POTENTIAL).
CC DOMAIN 1601 1664 COILED COIL (POTENTIAL).
CC SEQUENCE 1727 AA; 197858 MW; F820BF8D9C132644 CRC64;
SQ
Query Match 3.8%; Score 118; DB 1; Length 1727;
Best Local Similarity 19.9%; Pred. No. 4.7; Indels 118; Gaps 19;
Matches 104; Conservative 78; Mismatches 223;
Qy 34 DHPDSSRELSKILAKLERLQKQEDLRMAESLRIPGPDQGTATGRVRLVLEQLVKA 93
Db 1166 NHVKRISKEMETISQORQLLFLENKLRKTVESSNRVIADLQRG-----ITEKDVST 1218
Qy 94 KEQI---ENY-----KKQARNGLGKDHET---LRRRIENGAKELWFF----- 129
Db 1219 SESVGERSNYLNVALLNESKSLRENLERNEEVITELREKLETTLKTLDFANRLKEQLE 1278
Qy 130 ---LQSELKKLHLEGN---ELORHADAELLDLGHHSIRSIWTDLYLSQTDGADWEKE-- 183
Db 1279 SOLQTEKAAVKKLENSNEYKRRHQEILLNLN-----SSTSTSSDARLKNNE 1325
Qy 184 ---AKDLTELVRRTYILQNPKDCSKARKLVCNINKGCGYCOLHHVVVYCFMAYGTORT 240
Db 1326 LVSKENLIEELNQEIGHLKSELETVKSKSEDLNER-----AQNSKI 1368
Qy 241 LILESQNWRYATGGWETVFRPVSETCTDRSGLSGTGHSNVDNKNIQVLELPIVDSLHPR 300
Db 1369 EQLEKNTKLA-AAWRTKYEQVWVNSLEKHQIRQOLSKTSELEAKVAECHQLNEQLNK 1427
Qy 301 PPPLPLAVPELDLADRLRLVHGDPVWVVSQFVYLIRPQPWLEKEIEEATKLGFKHPVI 360
Db 1428 PSATPTATTQ-----SEPSTVSEBENS-----TKELSTQKLSLIMDL 1469
Qy 361 GVHVRRTDKV-----GTEAAFPH-PTEEYV-----VHVEEHFOLLARRMOVDKRVVLA 407
Db 1470 NTTKEELKVQNSKNKSGTSDKDTETPNEEMERKVMQOEVLRLRSIAKELQKNELLR 1529
Qy 408 TDDPTLLKEAK-----TKYSNYEFISDINSISWSAGLHNRVTENSLRGVILDIHF----- 456
Db 1530 KQNVQLQDVQVQALQETVVSSEAEASVHADTKDLENLKKTEEMLSVTFQVIFNESISDF 1589
Qy 457 -LSQADFLVCTFFSSQVCRVAYEIMQTLHPDASANPH--SLDDI 496
Db 1590 STSTADF--TFVQKEWEKREIRILLQKQVEEQVAQSHQQLDNI 1630

```

1052 -----RLEEDLADSRATSSA-----IENDIGNATGELRSSEHNA 1086

```

Db 522 LSLNQLBGLREL--GFQTPPE---DLAKAEQKRLMLRERLEBELRKLNSLEBKVRN-LS 575
Qy 110 KQHEILR-----RRI---ENGAKELWFFLOSELKKLKHLEGNELQORHADELILLD 156
Db 576 REVALREAKTRALEVLQRLGKKEEAREKLTLSSESKLERM-----LVSKAEADLATFL 631
Qy 157 GHHERSIMTDLV-----YLSQTD---GAGDWREKEAKOL-----TELQVRRITVYL 198
Db 632 GITAYRSLDLDLEKAREALEGVVDKLSATERRLEEARLKEEAALKWBAEQVMKRLEEL 691
Qy 199 QNPCKDSKARKLVNINCKGCGVCOHLHVYCFMAYGTCORTLI-LESQWRYATGGWET 257
Db 692 E-----AEBKLUKKEVSRKSEIARLKEY-----ONTLAELEDDRIISR----- 728
Qy 258 VFRPVSETCTDRSGLTGSHWSGEVNDKNTQVVELPIVDSLHPRPVPLPLA-----VPEDILA 313
Db 729 IDREMGELQTRIREMKSRKASE-----EALKL-----YLPAAASRRINEEIG 771
Qy 314 D-----RLLRVHGDPVAMVWSQFVKYLIRPOPWLEKEIEBATKLGFKHPVIGVHVR----- 365
Db 772 ETAYRELLAV-----LEDEMNDILSR--FNLDVAGVEIREKAA 807
Qy 366 ---RTDKVGTAEAFPHIEYMHVHEHFOL-----LARRMQVDKRVYLATDDPTLLKEA 417
Db 808 REIEVKAIGGNGAYRPLE--AVSGGERTVLALSFVLALNKAVGGKLGFLALDEPTANLDE 865
Qy 418 KTKYSNVEFISDINSWSAGLHNRYSLSLGRVILDIHFLSQADF--LVC--TFGSQVCR 473
Db 866 DRRRLSLVEVLRGISV-----EGLVRQLVVVTHEDVRDVTCLVTRIQGSR 914
Qy 474 V 474
Db 915 V 915

RESULT 13
MYSAL CAEEL
ID MYSAL CAEEL STANDARD; PRT; 1969 AA.
AC P12824; Q21440.
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myosin heavy chain A (MHC A).
GN MYO-3 OR K12F2.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
ON NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=89178677; PubMed=2926820;
RA Dibb N.J., Maruyama I.N., Krause M., Karn J.;
RT "Sequence analysis of the complete Caenorhabditis elegans myosin
RT heavy chain gene family.";
RL J. Mol. Biol. 205:603-613(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Harris B.R.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- FTM: TWO CYSTINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE

```

```

CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -!- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN
CC C.ELEGANS.
CC -!- MISCELLANEOUS: MHC A AND MHC B ARE FOUND EXCLUSIVELY IN THE BODY
CC WALL MUSCLE. THEY CO-ASSEMBLE INTO BODY WALL THICK FILAMENT.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X08067; CAA30356.1; -.
CC EMBL; Z78199; CAB01576.1; -.
CC PIR; T23622; S02771.
CC HSP; P08799; 1MND.
CC WormPep; K12F2.1; C312204.
CC InterPro; IPR000048; IQ_region.
CC InterPro; IPR001609; myosin_head.
CC InterPro; IPR004009; Myosin_N.
CC InterPro; IPR002928; Myosin_tail.
CC Pfam; PF00063; myosin_head; 1.
CC Pfam; PF01576; Myosin_tail; 1.
CC PRINTS; PR00193; MYOSINHEAVY.
CC ProDom; PD000355; myosin_head; 1.
CC SMART; SM00015; IQ; 1.
CC SMART; SM00242; MYSC; 1.
CC PROSITE; PS50096; IQ; 1.
CC Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
CC ATP-binding; Methylation; Alkylation; Multigene family.
KW MYOSIN HEAD-LIKE.
FT DOMAIN 1 793
FT DOMAIN 794 923
FT DOMAIN 857 1969
FT NP_BIND 179 186
FT ACTIN_BINDING.
FT DOMAIN 667 589
FT DOMAIN 770 784
FT MOD_RES 130 130
FT MOD_RES 707 707
FT MOD_RES 717 717
FT CONFLICT 116 116
FT CONFLICT 116 116
FT CONFLICT 116 116
SQ SEQUENCE 1969 AA; 225509 MW; 64577BBAF7EAD80A CRC64;

Query Match 3.8%; Score 116.5; DB 1; Length 1969;
Best Local Similarity 19.1%; Pred. No. 7.1;
Matches 89; Conservative 69; Mismatches 148; Indels 161; Gaps 19;

Qy 42 ELSKILAKLERLKQONEDLRMAE-----SLRPEGPDQGTATGRVTV 85
Db 931 KLSIDITQLEDMQERNEDLARQKKTQDELSDTKKHVQDLESLRKAQEQKSRDHNRS 990
Qy 86 LEEQLVKAKEQIENYKQARNGLGKHDEILRRRIENGAKELWFFLOSELKKLHLE--GN 143
Db 991 LQDEMANQDEAVAKLNKE-----KKHQ-----EESNRKLNEDLQSEDKVNHLEKIRN 1038
Qy 144 ELQORHADEILLDLGHHERSIMTDLVYLSQTDGAGDWREKEAKDLTELQVRRITVYLQNPXD 203
Db 1039 KLEQOMDEL-----EENID-----REKSRGDIKAKRKV-----EGD 1071
Qy 204 CSKARKLVNINCKGCGVCOHLHVYCFMAYGTCORTLI-LESQWRYATGGWETFRPVS 263
Db 1072 LKVAQENIDEITK-----QKHV-----
Qy 264 ETCIDRSGLSTGSHWSGEVNDKNTQVVELP-IVDSLHPRPVPLPLAVPEDL-ADRLLRVHG 321
Db 1090 ETLTKREEDLHHTNKLAEANSIIIAKLQRLKELTARNAL-----ELEEARNRQKS 1145
Qy 322 DPAVWVWSQFVKYLIRPOPWLEKEIEBATKLGFKHPVIGVH-----R 365

```



Db 1146 D-----RSRSEARELEELTERLEQGGGATAAQLEANKKREAIKLRREK 1191  
Qy 366 RTDKVGTAEAFHPIE-----EYVHVVEHFOLLAR-----RMQVDKRVYLATDDP 411  
Db 1192 EEDSLNHETAISSLRKRGDSVAELTEQLETLQKAKSEAKSKLORDLEESQATDSE 1251  
Qy 412 T-----LLKEATKYNYRPFISD---NSISWSAGLHRY-TENS 446  
Db 1252 VRSRQDLEKALKTIEVOYSELQTKADEQSRQLQDFAALKNRLNNS 1298  
RESULT 14  
REST HUMAN  
ID REST HUMAN STANDARD; PRT; 1427 AA.  
AC P30622;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Restin (Cytoplasmic linker protein-170 alpha-2) (CLIP-170) (Reed-  
Sternberg intermediate filament associated protein).  
GN RSN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Tissue=Peripheral blood monocytes;  
RX MEDLINE=92289675; PubMed=1600942;  
RA Bilbe G., Delabie J., Ruegg J., Richener H., Asselbergs F.A.M.,  
RA Carletti N., Sorg C., Odink K., Tarcsay L., Wiesendanger W.,  
RA de Wolf-Peters C., Shipman R.  
RT "Restin: a novel intermediate filament-associated protein highly  
RT expressed in the Reed-Sternberg cells of Hodgkin's disease."  
RL EMBO J. 11:2103-2113(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92405160; PubMed=1356075;  
RA Pierre P., Scheel J., Rickard J.E., Kreis T.E.;  
RT "CLIP-170 links endocytic vesicles to microtubules."  
RL Cell 70:987-900(1992).  
CC -1- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN  
CC THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE  
CC CYTOSKELETON.  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=Long;  
CC IsoId=P30622-1; Sequence=Displayed;  
CC Name=Short;  
CC IsoId=P30622-2; Sequence=VSP\_000765;  
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBERG CELLS  
CC OF HODGKIN'S DISEASE.  
CC -1- SIMILARITY: Contains 2 CAP-Gly domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; X64838; CAA46050.1; -;  
CC EMBL; M97501; AAA35693.1; -;  
CC PIR; S22695; S22695.  
CC Genew; HGNC:10461; RSN.  
CC MTM; 179838; -;  
CC GO; GO:0005768; C:endosome; TAS.  
CC GO; GO:0005882; C:intermediate filament; TAS.  
CC GO; GO:0015630; C:microtubule cytoskeleton; TAS.  
CC GO; GO:0008017; F:microtubule binding activity; TAS.

DR GO; GO:0006899; P:non-selective vesicle transport; TAS.  
DR InterPro; IPR000938; CAP-Gly.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF01302; CAP\_GLY\_2.  
DR SMART; SM00343; Znf\_C2HC; 1.  
DR PROSITE; PS00845; CAP\_GLY\_1; 2.  
DR PROSITE; PS02045; CAP\_GLY\_2; 2.  
KW Cytoskeleton; Microtubules; Coiled coil; Repeat; Alternative splicing.  
FT DOMAIN 78 120  
FT CAP-GLY 1.  
FT DOMAIN 143 204  
FT SER-RICH.  
FT DOMAIN 232 274  
FT CAP-GLY 2.  
FT DOMAIN 304 331  
FT SER-RICH.  
FT DOMAIN 350 1342  
FT COILED COIL (POTENTIAL).  
FT CCHC-BOX.  
FT MISSING (in isoform Short).  
FT /FTID=VSP\_000765.  
FT VARSPLIC 457 491  
FT D -> E (IN REF. 2).  
FT CONFLICT 1069 1069  
FT SEQUENCE 1427 AA; 160989 MW; 0A4F166DD94254E8 CRC64;  
SQ  
Query Match 3.7%; Score 115.5; DB 1; Length 1427;  
Best Local Similarity 17.9%; Pred. No. 5.5;  
Matches 101; Conservative 97; Mismatches 208; Indels 157; Gaps 20;  
Qy 40 SRELSKILAKLERLKKQNEDEL-----RRMAESLRIPGPIDQGTATGRVR 84  
Db 588 SKENESLKLEHANKENSDVIALWKSLETAIAHQOAMEELKVFSFK-GLGTETAFA 646  
Qy 85 VLEQLVKAK-----EQIENYKKQARNGIGKQHEILRRRIENGAKELWFFLOS----- 132  
Db 647 ELKQIEKWRLDYQHEIENLQNDQDSERAAHAKEMEARAKLMKVIKENSLEAIRSKL 706  
Qy 133 -----ELKKLKHLEG--NELORHADEILLDLGHHSIMTDL 167  
Db 707 DKAEDQHLVEMEDTLNKLQEAIEIKVKELEVLQAKNEQTKVIDNFTSQLKATEEKL-L 765  
Qy 168 YLSQTDGAGDWREKADLTVELYRRITTYLQNPK--DCSKARKL-----VCNIN 215  
Db 766 DALRKASSEGKSEMKKGRQQLAAEKQIKHLEIEKNAESSKASSITRELOGBELKLTNLQ 825  
Qy 216 KGCYGCQLHHVYVCFMAYGTQRTLLILESONWRYATGWTVPFVPSVETCTDRSGLSTG 275  
Db 826 ENLSEVSVQKETL-----EKELQILKEKFAEAEVSVQSMQET----- 866  
Qy 276 HWSGEVDKNTQVVELPVDLSLHPPVPL-----AVPEDLADRLLRVHGDPVWVVS 329  
Db 867 -----VNKLHQEQFNMLSSDLEKLRENLMADMEAKFREK-----D 902  
Qy 330 QVVKYLIRPQWLEKEIEEATK-----KLGFKHPVIGVHVVRTDKVGTAA 375  
Db 903 EREEQLIRAKLEKLENDIAEIMKMSGDNSSQLTKONDELRLKRDVEELQLKLTKANENAS 962  
Qy 376 F--HPIEEYVMVHEHFOLLARRMOVKRKYVYLATDDPTLLKEAKTKSYNEFFIS---DN 430  
Db 963 FLOKSIEDMTVKAESQQAQKHEEKLEKRLSD--LEKKMETSHNQOQLKARVER 1020  
Qy 431 S1SWSGALHRY-----TENSRL-----GVILDIHFL-SQADFLVCTSSQVC 472  
Db 1021 ATSETTKTHEELQLNQLTKLTDLTDKLGARENSGLLQLEELRKQAD---KAKAAQTA 1077  
Qy 473 RVAYEIMQTLHPDASANFHSLLDD 495  
Db 1078 EDAMQIMEQMTKEKTETTLASUED 1100  
RESULT 15  
OSFL\_HUMAN  
ID OSFL\_HUMAN STANDARD; PRT; 214 AA.  
AC Q92882;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Osteoclast stimulating factor 1.  
GN OSTFI.





**THIS PAGE BLANK (USPTO)**

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	2830.5	91.8	578	13	Q8AXS7	Q8axs7 xenopus lae
2	2187	70.9	446	4	Q8NEP2	Q8nep2 homo sapien
3	1480.5	48.0	619	5	Q9VYV5	Q9vvy5 d putative
4	978	31.7	559	5	Q16882	Q16882 caenorhabdi
5	810	26.3	169	4	Q8IUAS	Q8iuas homo sapien
6	670	21.7	512	5	Q8IFW9	Q8ifw9 ciona intes
7	601	19.5	560	5	Q8IFW8	Q8ifw8 ciona intes
8	130.5	4.2	1137	16	Q8F3E7	Q8f3e7 leptospira
9	129.5	4.2	1218	5	Q8IED2	Q8ied2 plasmodium
10	126	4.1	876	17	Q8TXI4	Q8txi4 methanopyru
11	125	4.1	1025	10	Q9SAA5	Q9saas arabidopsis
12	122	4.0	745	11	F70565	F70565 rattus norv
13	121	3.9	462	16	Q8RIM2	Q8rim2 fusobacteri
14	120.5	3.9	3542	5	Q9U5M2	Q9u5m2 plasmodium
15	120	3.9	1177	16	Q8RCY8	Q8rcy8 thermocoaer
16	119.5	3.9	681	11	Q8BU18	Q8bu18 mus musculu

```
Db 121 ATEANGAFYLVQSEVKKLKHLDNELQHVDEILIDMGHOORSVMTDLYLVSQTDGAG 180
QY 178 DMREREAKDLTELVRRTYLYLQNPDCSKARKLVNINKGCGYGCQLHHVVCYCFMAYGT 237
Db 181 DMREREAKDLTELVRRTYLYLQNPDCSKARKLVNINKGCGYGCQLHHVVCYCFMAYGT 240
QY 238 QRTLLLESQWRVATGNETVPRVSETCTDRSGSLSTGHSSEVNDKNOVVELPIVDSL 297
Db 241 QRTLLLESQWRVATGNETVPRVSETCTDRSGSLSTGHSSEVNDKNOVVELPIVDSL 300
QY 298 HRPBPPLPLAVPDLADRLRLRHGDPVAVWVSQFVKYLIRPQWLEKEIEEATKGLPKFH 357
Db 301 HRPBPPLPLAVPDLADRLRLRHGDPVAVWVSQFVKYLIRPQWLEKEIEEATKGLPKFH 360
QY 358 PVIGVHVRTDKVGTAAFPHEEYMHVVEEHFQLLARMQVDKRVYLATDPTLLKEA 417
Db 361 PVIGVHVRTDKVGTAAFPHEEYMHVVEEHFQLLARMQVDKRVYLATDPTLLKEA 420
QY 418 KTKYSNYEFISNISWSAGLNRYTENSIRGVILDIHFLSQADFLVCTFSSQVCRVAYE 477
Db 421 KAKYPOYEFISNISWSAGLNRYTENSIRGVILDIHFLSQANFLVCTFSSQVCRVAYE 480
QY 478 IMQTLHPDASNFSLDDIYFVGQNAHNOIAVYHPKPRTEEEIPMEPGDIIIGVAGNHW 537
Db 481 IMQTLHPDASNFSLDDIYFVGQNAHNOIAVYHPKPRTEEEIPMEPGDIIIGVAGNHW 540
QY 538 GYSKGINRKLGTGLTGLPSYKVKREKTIETVYKPYPEAEK 575
Db 541 GYSKGINRKLGTGLTGLPSYKVKREKTIETVYKPYPEAEK 578

RESULT 2
Q8NEP2 PRELIMINARY; PRT; 446 AA.
AC Q8NEP2;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DE Putative alpha-(1,6)-fucosyltransferase (EC 2.4.1.68) (Glycoprotein 6-
DE alpha-L-fucosyltransferase) (GDP-fucose-glycoprotein
DE fucosyltransferase) (GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alpha1,6-
DE fucosyltransferase) (Alpha1-6FucT).
GN FUCT6 OR CG2448.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Petit D., Picaud F., Dupuy F., Germot A., Julien R., Maftah A.;
RT "Core a3- and a6-fucosyltransferases in Drosophila: characterization
RT and origin of diversity.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA STRAINS=Berkeley;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeller K., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milghina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
```

```
QY 161 RSTMTDLYLSOTDAGDREKAKDLTELVRRTYLYLQNPDCSKARKLVNINKGCGY 220
Db 32 KSMTDLYLSOTDAGDREKAKDLTELVRRTYLYLQNPDCSKARKLVNINKGCGY 91
QY 221 GQQLHHVVCYCFMAYGTQRTLLLESQWRVATGNETVPRVSETCTDRSGSLSTGHSSE 280
Db 92 GQQLHHVVCYCFMAYGTQRTLLLESQWRVATGNETVPRVSETCTDRSGSLSTGHSSE 151
QY 281 VNDKNTQVVELPIVDSLHPRPPLPLAVPDLADRLRLRHGDPVAVWVSQFVKYLIRPQ 340
Db 152 VKDKNQVVEFFIVDSLHPRPPLPLAVPDLADRLRLRHGDPVAVWVSQFVKYLIRPQ 211
QY 341 WLEKEIEEATKGLPKFHGHPVIGVHVRTDKVGTAAFPHEEYMHVVEEHFQLLARMQVD 400
Db 212 WLEKEIEEATKGLPKFHGHPVIGVHVRTDKVGTAAFPHEEYMHVVEEHFQLLARMQVD 271
```

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley; TISSUE=Testis;  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J., Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R., Gonzalez M., Guatin H., Li P., Liao G., Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M., Celniker S.;  
 RA Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 RL [4]  
 RN HOMOLOG.  
 RP PubMed=11698403;  
 RA Roos C., Kolmer M., Mattila P., Renkonen R.;  
 RA "Composition of *Drosophila melanogaster* proteome involved in fucosylated glycan metabolism".  
 RT J. Biol. Chem. 277:3168-3175(2002).  
 RL -!- FUNCTION: CATALYZES THE ADDITION OF FUCOSE IN ALPHA 1-6 LINKAGE TO THE FIRST GLCNAC RESIDUE, NEXT TO THE PEPTIDE CHAINS IN N-GLYCANS (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: GDP-L-FUCOSE + N4-(N-ACETYL-BETA-D-MANNOSYL)-1,2-ALPHA-D-MANNOSYL-1,3-(R(1)-ALPHA-1,6)-BETA-D-MANNOSYL-BETA-N-ACETYL-1,4-D-GLUCOSAMINYL-1,4-N-ACETYL-D-GLUCOSAMINYL)ASPARAGINE = GDP + N4-(N-ACETYL-BETA-D-GLUCOSAMINYL-1,2-ALPHA-D-MANNOSYL-1,3-(R(1)-ALPHA-1,6)-BETA-D-MANNOSYL-1,4-BETA-N-ACETYL-D-GLUCOSAMINYL)-1,4-(ALPHA-L-FUCOSYL-1,6)-N-ACETYL-D-GLUCOSAMINYL)ASPARAGINE.  
 CC -!- PATHWAY: GLYCOSYLATION.  
 CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND FORM IN TRANS CISTERNAE OF GOLGI (BY SIMILARITY).  
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 23.  
 DR EMBL: AF41264; AAK63649.1; -  
 DR EMBL: AF003487; AAF48079.1; -  
 DR EMBL: AY051451; AAK2875.1; -  
 DR FlyBase: FBgn0030327; CG2448.  
 DR InterPro: IPR001452; SH3.  
 DR Pfam: PF00018; SH3.  
 DR SMART: SM00326; SH3; 1.  
 DR Hypothetical protein; Transferase; Glycosyltransferase; Transmembrane; Signal-anchor; Golgi stack; SH3 domain.  
 KW Domain  
 FT DOMAIN 1 17 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 18 38 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).  
 FT DOMAIN 39 619 LUMENAL, CATALYTIC (POTENTIAL).  
 FT DOMAIN 548 609 SH3.  
 FT SITE 345 351 SH3-BINDING (POTENTIAL).  
 FT SITE 411 412 IMPORTANT FOR DONOR SUBSTRATE BINDING (BY SIMILARITY).  
 SQ SEQUENCE 619 AA; 70205 MW; 691BFD9B5C6557DE CRC64;  
 Query Match 48.0%; Score 1480.5; DB 5; Length 619;  
 Best Local Similarity 46.5%; Pred. No. 3.7e-102;  
 Matches 285; Conservative 108; Mismatches 169; Indels 51; Gaps 8;  
 QY 3 AWTGSW-RWIMLLPAWGLTLLFYIGHLVRDN----DHPDSSRELKSLAKLERLKKQ 56  
 DB 10 ASANSWARALITIFLWAGLGVVFKLTNTGQQAGSELSNARRISQALQWLEHTROR 69  
 QY 57 NEDLRMAESLRIPEDIGQGTATGRVRLVEEQLVKAKEQIE----- 98  
 DB 70 NEELKQLIDELMSDQ--LQKQAMKLVQRLDNLNPKLPAVEAGPEPESFESAPADLR 127  
 QY 99 ----NYKKQARNGI-----GKQHEILRRRIENGAKELWFLQSELKKLK----- 138  
 DB 128 GWNVAEGAPNDLEAGVDPDHGFPSLEYEFTTRITQTNIGTWIWNFFSELGKVRKAVAA 187  
 QY 139 -HLEGNEQLQHADEILLDLGHHSRSMITDLYVLSQTDGAGDWREKAKDLTELVRITY 197

Db CHASA-DLEESINOVLLQGAHKKSLSDMERMRQSDGYEARHKEARDLSDLVORRLHH 246  
 QY LONPKDCSKARKLVNKNKGCGYCOLHHVYCFWIAVGTORTLILESQNNRYATGGWET 257  
 Db LQNPFDQONARKLVCKLNKGCGYCOLHHVYCFIVATATERTLLKSRGWRHYHKGWEE 306  
 QY VFRPVSETCTDRSGLSTGHSGEVNDKNIQVVELPIVDLSLHPRPPYLPVLPEDIALRL 317  
 Db VFQVSNNSCHDAGTANTYWFEGK--PNTQVLVLPIDSLMPRPYLPVLPEDIALRLK 363  
 QY RVHGDPVWVVSQVFKYLIRPQWLEKEIEBATKGLGPKHPIGVHVHRTDKVGTAAAFH 377  
 Db RLHGDPVWVVSQVFKYLIRPQWLEKEIEBATKGLGPKHPIGVHVHRTDKVGTAAAFH 423  
 QY PLEEVVHVEHFQOLLARRMQVDKRVYLATDDPTLLKEAKTKYSNYEFISDSWSAG 437  
 Db SVEEYTYVEDYRTLEVNGSTVARIPLASDDAQVIEBARKYPQYQIIGDPEVARWAS 483  
 QY LHNRYTENSJRGVILDIHFLSQADPLVCTFSQVCRVAYEIMQTLHPDASANFHSDDIY 497  
 Db VSTRYDTALANGIILDIHLLSMSDLVCTFSQVCRVAYEIMQTLHPDAAHRFKSLDDIY 543  
 QY YFGQNAHNQIAVYPHKKPTEEEIPMEPDIGIAGNHWGDSYKINRKLKGLTGLYPSYK 557  
 Db YFGQNAHNRRVIAHKPRTHEDILQLRVGDVSVAGNHWGDSYKINRKLKGLTGLYPSYK 603  
 QY VREKIEYVKKPTY 570  
 Db VEEKYDTAKPL 616

RESULT 4  
 OL6882 PRELIMINARY; PRT; 559 AA.  
 ID OL6882;  
 AC OL6882;  
 DT 01-JAN-1998 (Tremblrel. 05, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Hypothetical protein C10F3.6 (Core alpha-6-fucosyltransferase) (EC 2.4.1.68).  
 GN C10F3.6 OR FUT8.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_taxid=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA Waterston R.;  
 RT "Genome sequence of the nematode *C. elegans*: a platform for investigating biology. The *C. elegans* Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Davidson S., Wohlmann P.;  
 RT "The sequence of *C. elegans* cosmid C10F3.";  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Balazino L., Oriol R., Michalski J.C., Candellier J.J., Martinez-Duncker I., Mollicone R.;  
 RT "Cloning, expression and genomic organization of two core fucosyltransferases (Ce and CeD) from *Caenorhabditis elegans*.";  
 RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF022968; AAN84870.1; -





RA Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,  
 RA Berriaman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,  
 RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.,  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL844509; CAD52328.1; -  
 SQ SEQUENCE 1218 AA; 143059 MW; EIDBD78388BC7ESB CRC64;

Query Match 4.2%; Score 129.5; DB 5; Length 1218;  
 Best Local Similarity 18.4%; Pred. No. 1;  
 Matches 87; Conservative 88; Mismatches 157; Indels 141; Gaps 18;

QY 32 DNDPHSSRELKILAKLERLQKQEDLRMAESLRIPGPDQ-----GTATG 81  
 DB 707 DNE-----KLKESEKLSLKEAEKKIISIKELQIYENNLINMETSKYGVNK 760  
 QY 82 RVRVLEOLVKAEOI-ENYKQAR-----NGLGKD---HEILRRRIENGAKELWFFLOS 132  
 DB 761 KIEHKNEDKGNELSELYKEQKLTETVRKLEKIDISEYANKOKKEEDLKEIKKLQN 820  
 QY 133 ELKKLHLEGNELQRLHADEILLDLGHHSIMTDLYLSQTDGAGDWREKAKDLTE--- 189  
 DB 821 KIQLE-TEHKKKEEIDVLLQIENYKQKKEETNDLSSTDEIINEIEKKIEDIEKNIN 879  
 QY 190 LVORRTYLPONKDCSKARKLVNINKGCGYCOLHHVYVCFMAYGTORTLILESQNR 249  
 DB 880 ITKENLKELEN-----KITELQSFSS---SYENEMKHVVKKI----- 913  
 QY 250 YATGGWETVPRPVSETCTDRSGLSTGHWSGEVNDKNIQVVELPIVDSLHPRPPYLP 309  
 DB 914 -----EDLEKKSENLID-----LKKLENTLID-----LQ 938  
 QY 310 EDLADRLRLRHGDPVWVWSQFYKLYRPOPWLEKEIEBATKKL-----GFKHPVIGVH 363  
 DB 939 KDLKTS-----SDTVKLYKTHVWIESVEPLFNKKYTPYDFENFRHVIQKK 985  
 QY 364 VRTDKVGTAAAPHPPIEYVWVVEHFO-----LLARMQV---DKRV----- 404  
 DB 986 IQALONEQNKLIN-INRKAVQMYEQVDYKDLVTKSQVEBKKKIQEVIALDLVKK 1044  
 QY 405 -----YLATDDPTLLKEAKTKSYNYEFISDINSISWSAGLHNRYTEN 445  
 DB 1045 ESSLAMVQINQYFQAFSTFLLNNAQKLSIVDGLDGLANGIEMKIAPNNWNKES 1097

## RESULT 10

Q8TX14  
 ID Q8TX14 PRELIMINARY; PRT; 876 AA.  
 AC Q8TX14  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE SWC1-family ATPase involved in DNA repair.  
 GN SECC OR MK0690.  
 OS Methanopyrus kandleri.  
 OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;  
 OC Methanopyrus.  
 OC NCBI\_TaxID=2320;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AV19 / DSM 6324 / JCM 9639;  
 RX MEDLINE=21927647; PubMed=11930014;  
 RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,  
 RA Shcherbina O.V., Shakhova V.V., Belova G.I., Aravind L.,  
 RA Natalie D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,  
 RA Malykh A.G., Koonin E.V., Kozayavkin S.A.;  
 RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19  
 RT and monophyly of archaeal methanogens."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649 (2002).  
 RL EMBL: AE010362; AAC01904.1; -  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_Transporter.  
 DR InterPro; IPR003405; SMC\_C.  
 DR InterPro; IPR003395; SMC\_N.

## RESULT 11

Q9SAA5  
 ID Q9SAA5 PRELIMINARY; PRT; 1025 AA.  
 AC Q9SAA5  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE F25C20.13 protein.  
 GN F25C20.13  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosida II; Brassicales; Brassicaceae; Arabidopsia.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,  
 RA Li J., Lee J.M., Kremenetskaia I., Luros J., Ngan I., Liu A.,  
 RA Gonzalez A., Altafi H., Araujo R., Chao C., Conn L., Conway A.B.,  
 RA Dunn P., Hansen N., Huizar L., Kim C., Palm C., Rowley D., Shinn P.,  
 RA Walker M., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;  
 RT "Arabidopsis thaliana chromosome 1 BAC F25C20 sequence."  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Theologis A.;  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

DR InterPro; IPR002017; Spectrin.  
 DR Pfam; PF02483; SMC\_C; 1.  
 DR Pfam; PF02463; SMC\_N; 1.  
 DR SMART; SM00382; AAA; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 876 AA; 103706 MW; D651B682271FD383 CRC64;

Query Match 4.1%; Score 126; DB 17; Length 876;  
 Best Local Similarity 21.1%; Pred. No. 1.2;  
 Matches 100; Conservative 77; Mismatches 160; Indels 136; Gaps 21;

QY 42 ELSKILAKLERLQKQEDLRMAESLRIPGPDQ-GTATGRVVRVLEOLVKAKEIQENY 100  
 DB 245 EURLLENKIESLUGRRDDRLKLVGEGKAERLQRLQDVFPVSKRELENEAEELRRRIE-- 302  
 QY 101 KQKANGLGKDHIEILRRRIENGAKELWFFLQSELKKLKHLEG----- 142  
 DB 303 --ELRN-LLDDLRLSLRNRLSAAEEL-EGVKEBELSELKDEAGVDPERLVEFPKDIVASE 358  
 QY 143 -----NELQHADEI---LLDLGHHSIMTDLYLSQ---TDGAGDWREKAKDLTE 189  
 DB 359 RLRDLRREBELKRLKLEKVSDELSELGDRDEETLQSEYEELQERLDEIQGELKEIRVKE-KE 417  
 QY 190 LVORRTYLPONKDC-----SKARKLVNINKGCGYCOLHHVYVCFMAYGTORT 240  
 DB 418 LLERIESLREAGGECVCLRLPRERAELKLRDAEK-----ELERL-----QGREED 464  
 QY 241 LILESQNW-----RVATGGWETVFR-----PVSETCTDRSGLSTGHWSG 279  
 DB 465 LKERRELKORLESVRELEGTKEKRWRLRRERRELEBELKEELADLS----- 518  
 QY 280 EYNDKNIQVVELPIVDSLHPRPPYLPVLPEDLADRLRLRHGDPVWVWSQFYKLYRPOP 339  
 DB 519 --RELGVEDRLPELRLAVR-----AESLLRDLERRRGD-----VLRLEKELERTL 563  
 QY 340 PWLEK-----EIEATKKLGFKHPVIGVHVHRTDKVGTAAAPHPPIEYVWVVEH 389  
 DB 564 DRCEVIGTSPGSDVEELRLEERDHHVQKLEAE--GELERYHNLHEKVKRARE 621  
 QY 390 FOLLARRQVDRKRVYLATDDPTLLKEAKTKSYNYEFISDINSISWSAGLHNRY 442  
 DB 622 RKEL-KRIERD-----LEDAGRLQEQUERNLE-----GLRERY 653









**THIS PAGE BLANK (USPTO)**